Exploring the synthetic possibilities and siRNA delivery potential of Small Molecule Carriers (SMoCs)

by

Matthew John Gooding

Submitted in accordance with the requirements

for the degree of

Doctor of Philosophy

AUGUST 2011

UCL

I, Matthew John Gooding, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Abstract

Delivery of proteins and nucleic acids into cells is a major challenge to the development of biological therapeutics. Cell penetrating peptides (CPPs) and cationic liposomes have been shown to internalise short interfering RNA (siRNA) to achieve gene silencing, but no standard reagent exists which can safely deliver macromolecules both *in vitro* and *in vivo*. Small Molecule Carriers (SMoCs) are amphipathic α-helix mimetics displaying guanidine groups in order to mimic the structure of the CPP penetratin. Previously, SMoCs were shown to effectively deliver active proteins into cells. It is hypothesized that SMoCs may also be applied to the delivery of siRNA into cells in order to knockdown target genes. In addition, since cell surface binding is thought to be a crucial step in CPP internalisation, a new SMoC which optimises binding to proteoglycans may be more efficiently taken up. The aims of this thesis are to optimise the synthesis of the SMoCs in order to increase the quantity of product; to demonstrate that SMoCs may be used as siRNA delivery agents; and to design and synthesize a new SMoC which maximises siRNA uptake. The synthesis of SMoCs has been significantly enhanced, with the development of new reagents to improve the yields and cost of production. The electrostatic interactions of SMoCs with siRNA have been characterised, using NMR to examine a π-cation interaction which may contribute to anion binding, as well as determination of binding affinities using ITC and gel shift assays. Initial experiments using SMoC-siRNA complexes show significant mRNA knowdown which demonstrates the potential of SMoCs as siRNA delivery vectors. Finally, a new SMoC has been designed and synthesized which represents the first in a new class of dendritic SMoCs which are designed to maximise binding to the cell surface. This SMoC is also capable of delivering siRNA into cells, and may also be expanded by the addition of targeting peptides.

Table of Contents

List of Figures

List of Schemes

List of Tables

List of Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
Boc	Tert-butoxycarbonyl
bp	Base Pairs
Cbz	Carboxybenzyl
cDNA	Complementary DNA
CI	Chemical Ionisation
COSY	Correlation Spectroscopy
CV	Column Volumes
DCM	Dichloromethane
DIEA	*N*,*N*-Diisopropylethylamine
DMA	Dimethylacetamide
DMEM	Dulbecco's Modified Eagle's Medium
DMF	Dimethylformamide
DNA	Deoxyribonucleic Acid
dppf	1,1'-Bis(diphenylphosphino)ferrocene
EC_{50}	Half maximal effective concentration
EGFP	Enhanced Green Fluorescent Protein
EI	Electron Ionisation
ES	Electrospray Ionisation
FCS	Fetal Calf Serum
FDA	US Food and Drug Administration
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GFP	Green Fluorescent Protein
HATU	2-(1*H*-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate methanaminium
HIV	Human Immunodeficiency Virus
HRMS	High Resolution Mass Spectroscopy
LC-MS	Liquid Chromatography-Mass Spectrometer
LRMS	Low Resolution Mass Spectroscopy
m/z	Mass to charge ratio
MALDI	Matrix-Assisted Laser Desorption/Ionisation
mRNA	Messenger RNA
Ms-	Mesyl
MW	Microwave
NMP	*N*-Methyl-2-pyrrolidone
NMR	Nuclear Magnetic Resonance
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
ppm	Parts per million
RISC	RNA-Induced Silencing Complex

RNA	Ribonucleic Acid
RNase	Ribonuclease
rt	Room temperature
TFA	Trifluoroacetic Acid
THF	Tetrahydrofuran
TIPS	Triisopropylsilane
tlc	Thin layer chromatography
TMEDA	Tetramethylethylenediamine
TMSCl	Trimethylsilyl chloride
TOF	Time of Flight
UV	Ultraviolet

Acknowledgements

I would like to thank my supervisor, David Selwood, for his excellent guidance, support and constructive criticism over the last four years, and for allowing me to work on this fascinating project. In addition, many thanks go to all members of the group, both past and present, including Edith Chan, Henry Dube, Paul Gane, Cristina Posada, Filipa Quinteiro, Michela Simone and Roberta Worthington, for their advice and for their considerable wealth of scientific knowledge which has enabled me to complete the experiments detailed here. I would also like to thank the many members of the other labs who have helped me to complete this work, including the groups of Kai Stoeber and Gareth Williams, for assistance with the molecular biology experiments, and Tina Divita and Matt Webster of the ISMB Biophysics Centre at Birkbeck University for assistance with the ITC work. Special thanks also goes to Neale Foxwell at WIBR for his invaluable training and assistance in cell culture, and for the excellent support that he provides to the whole institute. Finally, I would like to thank my family for their support and understanding, and my friends, especially Paul Amies, for their wisdom and unyielding encouragement.

CHAPTER 1

Introduction

1. Introduction

1.1. Macromolecular Therapeutics

Over the last decade, there has been increasing interest in the field of macromolecular therapeutics, in which biological molecules such as proteins and nucleic acids, rather than traditional small molecule inhibitors, are used to treat a wide range of diseases. Macromolecular drugs are thought to hold the key to a new era of treatments that show greater complexity and higher specificity than current small molecule drugs. Proteins may be used to upregulate or restore function to defective biochemical pathways, or to introduce a novel function not normally present in the body[1] and nucleic acids may be able to silence mutated genes with high specificity, or replace non-functioning genes.[2] Small-molecule drugs, on the other hand, are limited in their ability to target many biological mechanisms such as protein-protein interactions and genes. They are also often abandoned during the early stages of development due to off-target side effects and toxicity problems.

The first protein therapeutic, human insulin, was approved by the FDA in 1982 to treat type I and type II diabetes mellitus by replacing the defective protein produced by individuals with the disease.[1] Since then, other protein therapeutics which restore function to a defective pathway, augment and existing pathway, or add a novel function to the body have been approved, including most recently etanercept (Enbrel®), a protein-based drug for treatment of rheumatoid arthritis, plaque psoriasis and other autoimmune diseases. Etanercept is a fusion protein combining human TNF receptor 2, which acts as a TNF inhibitor, with the Fc domain of human immunoglobulin G1 (IgG1) which increases the drug's serum half-life by binding to the salvage receptor FcRn.[3] In addition, the fastest-growing class of protein therapeutics are monoclonal antibodies (mAbs), which bind to defective cells and label them for destruction by the immune system.[3] The first mAb therapeutic to be approved was muromonab-CD3 in 1986, and was a murine antibody which targeted the T-cell CD3 receptor in order to suppress the immune response in organ transplant patients. Recent advances have enabled the production of human mAbs using transgenic mice possessing human IgG genes, or phage display libraries.[4] As a result, 24 mAb drugs have now been approved, with a further 30 in Phase 2/3 or

Phase 3 clinical trials at the end of 2010.[5] In addition, mAbs may be used to target drugs to their site of action by binding specific cell types, including conjugating antibodies to radionuclides to destroy tumour cells, such as tositumomab, a murine mAb linked to iodine-131 which is used to treat non-Hodgkin lymphoma. In total, more than 130 different protein therapeutics have been approved by the FDA,[1] with more in development, showing that protein therapeutics are becoming an increasingly important area of drug development, and with the growing wealth of crystal structure data, will continue to be so in future years.

Despite lagging some years behind protein therapeutics, oligonucleotides have also shown considerable potential for use in clinical treatments.[2] The first oligonucleotide drug candidates were based on antisense technology, whereby single stranded nucleic acid molecules would bind sequence specifically to their complementary mRNA target, thus triggering degradation of the duplex by the RNase H system. The first and only antisense drug to be approved by the FDA was fomiversin in 1998, which targets a main transcriptional protein of the cytomegalovirus (CMV) in patients with CMV retinitis, a disease leading to blindness common in AIDS sufferers.[2] Despite several other antisense-based therapies entering clinical trials, there have been no other drugs approved, mainly due to their limited efficacy. However, in the same year that the first antisense drug was approved, experiments by Fire *et al*.[6] paved the way for a new approach to oligonucleotide therapeutics which has since superseded the previous RNase H-based antisense technology. This was the discovery of the RNA interference (RNAi) pathway in which short interfering RNA (siRNA) or short hairpin RNA (shRNA) are able to catalytically silence mRNA transcription by binding to their complementary mRNA sequences and activating a previously unknown biochemical pathway leading to degradation of the mRNA. A number of siRNA based drugs have now entered clinical trials (Table 1), including ALN-RSV01 for treatment of the respiratory syncytial virus (RSV), which causes upper and lower respiratory tract infections.[7] The drug is now entering Phase IIb clinical trials, and so far has shown to be effective at reducing infection rates in preliminary studies (www.alnylam.com). Another siRNA-based drug undergoing Phase II clinical trials is PF-655 for

treatment of age-related macular degeneration (AMD) and diabetic macular edema (DME) and targets the gene RTP801 which is upregulated as a result of ischemia, hypoxia or oxidative stress.[8] These trials show the potential of oligonucleotide based therapeutics, and demonstrate their importance in the future of macromolecular drug technologies.

Table 1. Current progress of clinical trials of siRNA-based therapeutics.

Drug	Company/Institution	Disease	Target	Delivery System	Phase	Status
TD101	Pachyonychia Congenita Project	Pachyonychia congenita	Keratin K6a	Naked siRNA	Phase I	Completed
AGN211745	Allergan	AMD	VEGF receptor-1	Naked siRNA	Phase II	Terminated
Proteosome siRNA and tumour antigen RNA-transfected dendritic cells	Duke University	Melanoma	Immunoproteosome	Unknown	Phase I	Recruiting
SV40 vectors carrying siRNA	Hadassah Medical Organization	CML	Tyrosine kinase	Viral	Unknown	Completed
CALAA-1	Calando Pharmaceuticals	Refractory cancer	M2 subunit of ribonucleotide reductase	Transferrin-targeted cyclodextrin nanoparticle	Phase I	Recruiting
ATU027	Silence Therapeutics AG	Advanced solid cancer	Protein kinase N3	Cationic lipid lipoplexes	Phase I	Recruiting
QPI-1007	Quark Pharmaceuticals	Optic Nerve Atrophy	Caspase 2	Modified siRNA	Phase I	Recruiting
Bevasiranib	Opko Health, Inc.	DME/AMD	VEGF	Naked siRNA	Phase III	Withdrawn
PRO-040201	Tekmira Pharmaceuticals Corporation	Hypercholesterolemia	Apo-B	SNALP-siRNA	Phase I	Terminated
IL-10 siRNA	National Taiwan University Hospital	Preeclampsia	IL-10	Unknown	Unknown	Terminated

AHR siRNA	National Taiwan University Hospital	Neuroblastoma	Aromatic hydrocarbon receptor	Unknown	Unknown	Completed
siG12D LODER	Silenseed Ltd	Adenocarcinoma	KrasG12D	Biodegradable polymeric matrix	Phase I	Recruiting
I5NP	Quark Pharmaceuticals	Acute Kidney Injury, Delayed Graft Function	p53	Modified siRNA	Phase I/II	Recruiting
SYL040012	Sylentis, S.A.	Ocular Hypertension	Adrenergic receptor beta-2	Naked siRNA	Phase I/II	Recruiting
TBX3	University of California, Irvine	Cell Differentiation	TBX3	Unknown	Unknown	Unknown
ALN-VSP02	Alnylam Pharmaceuticals	Solid tumours	Kinesin spindle protein, VEGF	siRNA-SNALP	Phase I	Active
ALN-RSV01	Alnylam Pharmaceuticals	RSV	RSV nucleocapsid	Naked siRNA	Phase IIb	Recruiting
ALN-TTR01	Alnylam Pharmaceuticals	Transthyretin-mediated amyloidosis	Transthyretin	Unknown	Phase I	Recruiting

However, there are still many hurdles to be overcome before the macromolecular drug revolution is finally upon us. A recent study into the safety of therapeutics which are derived from a biological system, so-called biologics, including recombinant proteins, antibodies and vaccines, indicated that a higher proportion of these drugs have prompted safety alerts than for small molecule therapeutics.[9] Of the 174 biologic drugs that were approved in either the US or Europe between January 1995 and June 2007, 23.6% prompted regulatory action as a result of safety concerns compared with an average of 8% of small molecule drugs studied between 1975 and 1999. Other challenges that have continued to hinder the progress of macromolecular therapeutics include the control of biodistribution and degradation of macromolecules in the body, off-target effects and specificity, and delivery to intracellular compartments.[8] Many advances have been made that increase the stability of proteins and oligonucleotides *in vivo* resulting in a longer half-life in the body, including the chemical modification of oligonucleotide strands to prevent degradation by endogenous nucleases. Similarly, modification of proteins by the addition of polyethylene glycol (PEG) protects from proteases, as well as decreasing renal clearance.[1] Off-target effects are a particular issue for oligonucleotide drugs, including both traditional antisense oligonucleotides and newer RNAi-based agents. Despite these oligonucleotides being designed to target a single mRNA sequence, binding to additional genes is often observed using DNA arrays, including to sequences with little homology to the target.[10] Improved design of siRNA sequences using computational tools such as BLAST, as well as chemical modifications (Section 1.3.2.1) have reduced off-target binding.

Finally, one of the largest challenges to the development of protein and oligonucleotide-based drugs has been the lack of a suitable delivery agent to target these macromolecules to the intracellular machinery. Proteins and oligonucleotides are large and hydrophilic, thus preventing them from crossing the lipid membrane surrounding cells. Traditional small molecule drugs usually conform to Lipinski's rule of five,[11] which stipulates that molecules should be small and lipophilic, making them permeable to cell membranes. Over the last decade, much research has been

dedicated to finding an effective delivery mechanism for macromolecules, which have included viral carriers, liposomes and cell penetrating peptides (CPPs).[12]

1.2. Cell Penetrating Peptides

CPPs, also called protein transduction domains (PTDs), are short peptides, usually less than 30 amino acids long, which have the unusual property of being able to cross the cell membrane. In 1988 it was found that a transcription factor from HIV called the transactivator of transcription (TAT) was taken up by cells and migrated to the nucleus to activate transcription of the viral genome.[13] Shortly after, it was also found that the *Drosophila melanogaster* Antennapedia homeodomain was also internalised into cells.[14] These observations led to the identification of the first CPPs, the third helix of the Antennapedia homeodomain, named penetratin (RQIKIYFQNRRMKWKK),[15] and a short basic peptide derived from TAT consisting of amino acids 47-57 (RKKRRQRRR).[16] Many other CPPs have since been described, including chimeric peptides such as MPG, a fusion peptide derived from a hydrophobic domain from HIV gp41 and a hydrophilic domain from the nuclear localisation sequence of Simian vacuolating virus 40 (SV40) T-antigen.[17] CPPs can be divided into two main classes; the amphipathic CPPs, containing both a hydrophobic and a hydrophilic domain, such as TAT, penetratin, and MPG, and the polycationic CPPs containing mainly positively charged residues such as lysine and arginine. Some CPPs are covalently linked to their cargo, such as TAT which was incorporated into fusion proteins and successfully delivered to various mouse tissues.[18] Other CPPs are used to form stable non-covalent complexes with their cargo, such as Pep-1 which was designed to form complexes with proteins and peptides through hydrophobic interactions with a tryptophan-rich domain.[19] The formation of non-covalent complexes has also been particularly useful in the CPP-mediated delivery of oligonucleotides, which are highly negatively charged and therefore able to form electrostatic complexes with cationic peptides (Section 1.3.2.4). Over the last two decades, CPPs have been used to transport a wide range of high molecular weight cargoes into cells, including biologically active proteins

both *in vitro* and *in vivo*,[18, 20] quantum dots,[21] and oligonucleotides,[22-23] making them attractive candidates for delivery agents for macromolecular drugs.

1.2.1. Uptake Mechanism

One of the most challenging questions surrounding CPPs is the mechanism by which these peptides enter cells. Early studies using fluorescent dyes linked to CPPs concluded that internalisation was energy-independent, occurring at both 4 °C and 37 °C.[15, 24] However, this was later found to be an experimental artefact resulting from CPPs adhering strongly to the cell membrane and not removed under the washing conditions used in the early experiments.[25] When the cells were fixed with methanol for visualisation, the membranes were disrupted allowing the CPP-dye complex to enter the cells, thus giving false-positive results for CPP internalisation. In addition, the flow cytometry protocols did not distinguish between internalised fluorescent CPPs and those adhered to the cell surface. Methods used to detect internalised fluorescent CPPs now include a more thorough washing procedure, such as treatment with trypsin or heparin, in order to remove extracellular peptide. In experiments by Lundberg *et al.*,[26] Chinese Hamster Ovary (CHO) cells were incubated with various CPPs linked to GFP. These experiments used microscopy on live cells to determine the extent of cell uptake and showed that for all the CPPs tested, the bulk of the proteins were bound to the cell surface rather than internalised into the cells. Subsequent studies on various CPPs such as TAT, penetratin and polyarginine concluded that CPP internalisation is driven by endocytosis, and in particular lipid-raft macropinocytosis (Figure 1).[27-28] However, the uptake mechanism of CPPs is still unclear, and it is probable that there is no general mechanism that is applicable for all peptides and cargoes. Indeed, recent studies have suggested that CPP uptake takes place by both endocytosis and energy-independent translocation, with the balance between these pathways influenced by factors such as CPP sequence, temperature and CPP concentration.[29-30]

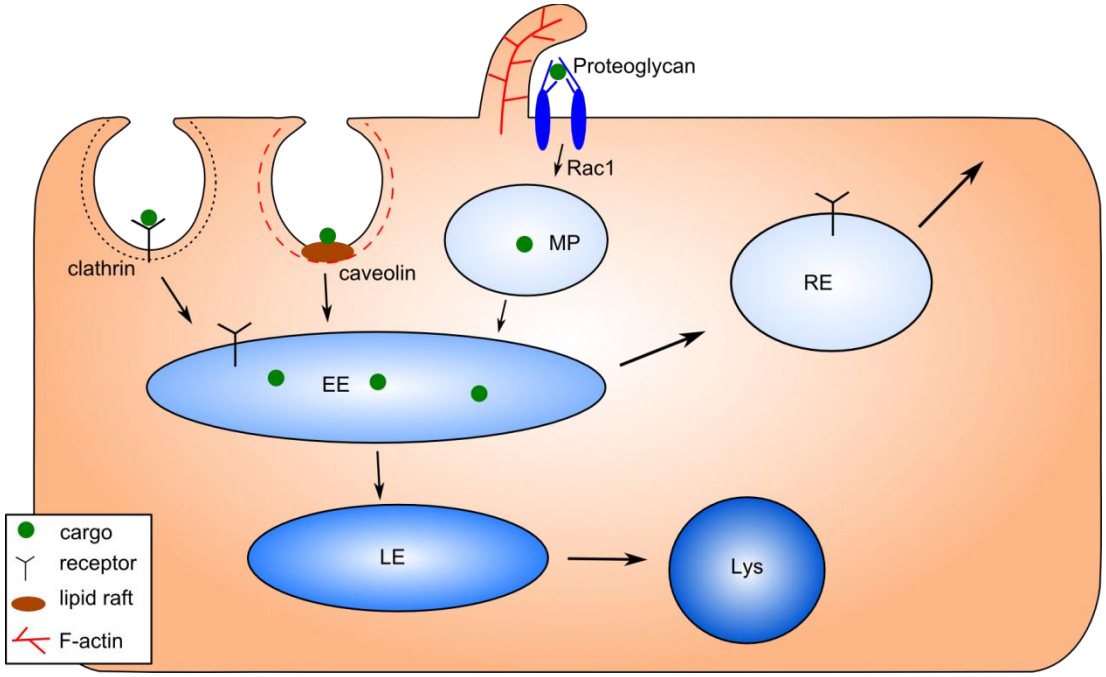


Figure 1. Endocytosis mechanisms used by CPPs. Clathrin-mediated endocytosis occurs at clathrin-coated pits which mature into clathrin-coated vesicles upon binding of a ligand to its cell-surface receptor. Dynamin releases the vesicles from the membrane allowing them to progress into the cell and fuse with early endosomes (EE). Some receptors, such as the low-density lipoprotein (LDL) receptor are returned to the cell surface via recycling endosomes (RE). Alternatively, caveolin-coated vesicles (caveolae) may form at lipid-rich regions of the membrane, so-called lipid rafts, which are also released by dynamin into the cytoplasm and fuse with EEs. Finally, larger cargoes may be internalised by macropinocytosis via interactions with proteoglycans in the extracellular matrix, which results in rearrangement of the actin cytoskeleton to form lamellipodia. These structures engulf the cargo, internalising it into macropinosomes (MP) which fuse to EEs. EEs mature into late endosomes (LE) and finally lysosomes (Lys) where the contents is broken down by enzymes. Therefore, it is important that the CPP-bound cargo escapes from early endosomes in order to exert its biological effect.

1.2.1.1. *Cell-surface binding*

Many studies have shown that CPPs bind strongly to the glycosaminoglycans (GAGs) of cell surface proteoglycans, such as heparan sulfate (HS) and chondroitin sulfate (CS) as a precursor to internalisation.[26, 28, 31-35] Proteoglycans such as syndecans form part of the extracellular matrix, where they play a structural role in cell adhesion, as well as a regulatory role in mediating lipid microdomains when

bound by heparin-binding ligands.[36] It is thought that proteoglycan ligand binding results in receptor clustering, leading to the reorganisation of the actin cytoskeleton via activation of a signalling cascade involving the Rho-family of GTPases, therefore leading to cellular uptake via macropinocytosis.[37]

HS a linear polysaccharide made up of highly sulfated NS domains and low sulfated NA domains. Heparin is a closely related compound that has been used as a substitute for HS in binding assays with a number of CPPs which have calculated dissociation constants in the low micromolar to low nanomolar range.[38] These strong interactions are probably comprised of an ionic contribution as well as bi-dentate hydrogen bonds between the guanidine groups of the peptide's arginine amino acids and the sulfate anions of heparin. The internalisation of nonaarginine (R9), which was found to bind heparin with a K_d of 109 nM, was found to be GAG-dependent by comparing uptake in wild-type CHO cells with mutant cells deficient in HS and CS.[34] In GAG-deficient CHO cells, oligoarginine and TAT internalisation is significantly reduced compared to wild-type cells and the presence of CPPs results in activation of Rac1, a Rho-family GTPase which is involved with F-actin reorganisation, suggesting that GAG binding may be a precursor for CPP uptake by endocytosis.[31] Similarly, it was found that the CPPs MPG-α and MPG-β, either alone or in complexes with DNA, initiated remodelling of the actin network, leading to the formation of lamellipodia, accompanied by upregulation of Rac1.[35] In addition, CPPs and other polycationic cell penetrating compounds (CPCs) are capable of clustering GAGs, which is a precursor for endocytosis.[39-40]

A recent study, however, contradicts these findings claiming that whilst TAT does bind to cell-surface HS, this interaction is independent of uptake, with enzymatic removal of cell surface GAGs having little effect on the uptake of fluorescently-tagged TAT.[41] Contradictions such as these are commonplace in the literature surrounding CPP internalisation, and demonstrates our limited understanding of the mechanism of uptake of these peptides. It is possible that competing internalisation

mechanisms exist and that the pathway used is dependent on factors such as CPP sequence, cargo and cell line.

Another possible route by which CPPs are internalised into cells involves binding to anionic lipids on the cell surface. It is thought that the positively charged side chains on the peptide interact with anionic membrane lipids, shielding the positive charge and forming a neutral, lipophilic complex that is able to cross the membrane.[40, 42-44] It has been shown that addition of anionic lipids such as 1,2-dioleoyl-*sn*-glycero-3-phosphatidylglycerol (DOPG) are capable of shifting basic peptides from the aqueous phase to the organic phase by the formation of lipophilic complexes.[42, 45] This suggests that cell-surface phospholipids may act as counter-ions to the CPP's positively charged amino acids, allowing it to dissolve in the hydrophobic cell membrane. It is suggested that arginine favours this counterion-mediated membrane translocation over lysine, due to the tendency of guanidinium groups to form bidentate hydrogen bonds with anions in order to reduce charge repulsion with adjacent arginine residues, whereas charge repulsion in lysines is generally countered by a reduction in pK_a (Figure 2).[45] This may explain the high proportion of arginine residues present in CPPs. In addition, the presence of a hydrophobic counteranion, pyrenebutyrate, increases the translocation of arginine-containing CPPs.[30] Polyarginine is the most affected, followed by TAT, and then by amphipathic CPPs, showing that a high arginine content favours binding to counterions. The addition of the counteranion increased membrane perturbation in large unilamellar vesicles (LUVs), showing that increased hydrophobicity of the CPPs leads to greater penetration of the membranes. In a splice-switching assay in which the same CPPs were used to transport an antisense oligonucleotide complexed electrostatically to the peptides into HeLa cells, pyrenebutyrate was found to increase the rate of translocation, and therefore the biological effect, for arginine-rich peptides more than amphipathic peptides. However, it was found that chloroquine, which delays the maturation of endosomes into lysosomes, was required for a biological effect to be observed, hence pointing to an endocytotic internalisation pathway, with

pyrenebutyrate enhancing translocation through the endosomal membrane, rather than direct membrane penetration.

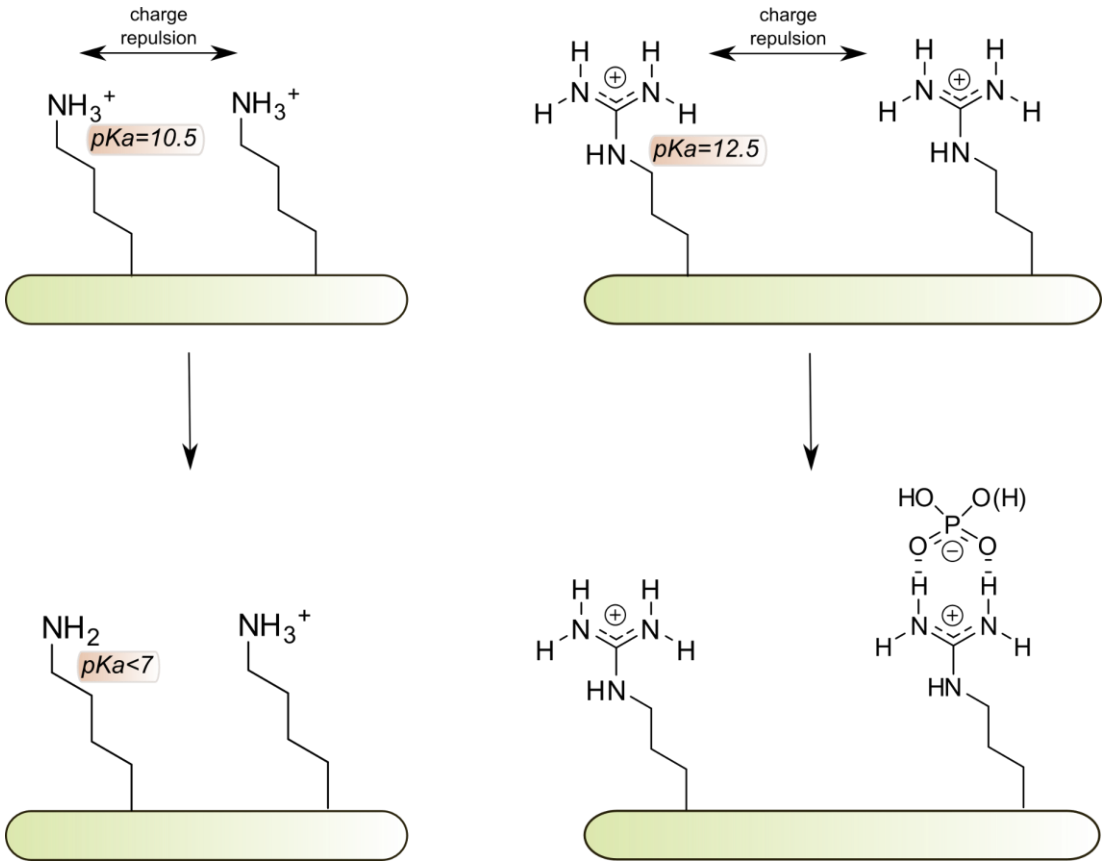


Figure 2. Reduction of charge repulsion by lysine and arginine oligomers. Amino groups in polylysine (left) reduce charge repulsion by lowering their pK_a, resulting in deprotonation. Guanidine groups in polyarginine (right) form bidentate hydrogen bonds to anions such as phosphates. Arginine-rich peptides may therefore more readily bind to counter-anions on the cell surface than lysine-containing peptides.

1.2.1.2. *Endosomal Escape*

Overcoming entrapment by endosomes is one of the major challenges to the design of efficient CPPs and other macromolecular transporters. By comparing the levels of endosomal escape of polyarginine- and polylysine-coupled liposomes, El-Sayed *et*

al. suggest that the reason for the improved endosomal escape of arginine-containing peptides over lysine-containing peptides lies in the ability of arginine to form bi-dentate hydrogen bonds with anionic membrane components (Figure 2).[44] These hydrogen bond interactions allow arginine residues to remain fully protonated on acidification of the endosome, therefore leading to increased potential across the membrane, whereas lysine residues may be deprotonated due to electrostatic repulsion. This increased membrane potential, together with the increased binding to the membrane, will therefore lead to a greater rate of endosomal escape than for lysine-containing peptides.

Biophysical studies have demonstrated that CPPs such as TAT and penetratin are capable of inducing negative Gaussian curvature in lipid membranes, which may explain the mechanism by which they are able to escape endosomes, or directly penetrate cell membranes.[46-47] Negative Gaussian curvature is an indicator of "saddle"-shaped membrane structures, which is a feature of membrane pores. Polylysine is only capable of inducing negative curvature in one direction along the membrane due to its electrostatic binding to anionic lipids. However, arginine is able to form bidentate hydrogen bonds with zwitterionic lipids, allowing it to induce positive curvature in the perpendicular direction to the negative curvature, resulting in negative Gaussian curvature. Although the mechanism of membrane interactions is not yet fully understood, and these results are derived from studies on synthetic membranes, it may help to explain the importance of arginine residues in CPPs. Another proposed mechanism for the endosomal escape of TAT suggests that the arginine residues do not interact directly with the membrane, but instead is involved in a pH-sensing mechanism that allows a conserved tryptophan residue to become exposed at low pH.[48] A triplet of three arginine residues are thought to interact with an N-terminal Glu residue, which becomes protonated in the acidic environment of the endosome, resulting in loss of this interaction resulting in a conformational change exposing the Trp side chain. Insertion of the Trp side chain into the membrane then results in endosomal escape. This mechanism explains the cell-penetrating ability of TAT and other amphipathic peptides, but not the cationic CPPs

which do not contain lipophilic amino acids. Therefore, endosomal escape may not follow a general mechanism and further supports the hypothesis that different CPPs are internalised via different pathways.

1.3. RNA Interference

1.3.1. Mechanism

In 1998, the Nobel prize-winning work carried out by Andrew Fire, Craig C. Mello and co-workers demonstrated that double stranded RNA (dsRNA) injected into the nematode worm *Caenorhabditis elegans* was capable of suppressing gene expression at a far higher potency than single stranded RNA (ssRNA).[6] The pathway by which dsRNA is able to silence gene expression is now known as RNAi and has been demonstrated in a wide range of eukaryotic organisms,[49-50] including humans.[51]

It was observed by Fire *et al*. that RNAi was unlikely to occur via a simple antisense mechanism due to the potency of the observed effects compared to ssRNA. Therefore, since its discovery, much attention has been given to the determination of the mechanism of the RNAi pathway (Figure 3). Hamilton and Baulcombe observed that when tomato plants were transformed with cDNA sequences of endogenous genes, gene silencing occurred, accompanied by the appearance of short 25-nucleotide RNA molecules complementary to the target gene.[52] The identification of a ribonuclease-III enzyme in *D. melanogaster* called Dicer (Dcr) provided an explanation for the presence of these short RNA strands.[53] Dcr is responsible for the 'initiation' phase of RNAi, processing long dsRNA strands into ~25 nucleotide siRNA molecules in an ATP-dependent mechanism.[53] *D. melanogaster* has been found to possess two Dcr enzymes, Dicer-1 (Dcr-1) and Dicer-2 (Dcr-2), of which Dcr-2 is the enzyme responsible for dsRNA cleavage.

The existence of RISC was elucidated around the same time as Dcr.[54] RISC is responsible for the 'effector' phase of RNAi and consists of a set of proteins which assemble around the siRNA strand. All RISC complexes so far isolated contain a protein from the Argonaute (Ago) protein family, such as Ago2 in *D.*

melanogaster,[55] possessing PAZ domains capable of binding 3' 2-nucleotide overhangs.[56] Following binding to RISC, one strand of the siRNA, the 'passenger' strand, is abandoned and the activated RISC-Ago2 complex together with the 'guide' strand identifies and binds to the complementary mRNA sequence, which is finally cleaved by endoribonuclease activity within the complex.[57] Proteins that link the initiation and effector phases, so-called dsRNA-binding proteins (dsRBPs), have been identified and include the *D. melanogaster* protein R2D2, so named because of its two RNA-binding domains (R2) and its ability to bind Dcr-2.[58] R2D2 does not affect the dsRNA-processing function of Dcr, however it has been shown to restore RISC function in a system containing purified Dcr and RISC, suggesting that its function is involved with loading siRNA onto RISC.[58] Human variants of R2D2 are yet to be elucidated, however two proteins, transactivating response RNA-binding protein (TRBP) and protein activator of protein kinase R (PACT), have been shown to bind to Dcr but their interaction with RISC is unclear.[59]

A mechanistic aspect of RNAi that is poorly understood is the separation of the two siRNA strands, resulting in just the antisense strand which guides RISC to its target mRNA sequence. Several proteins associated with RISC have been shown to possess helicase domains, but no direct link with the unwinding of siRNA strands has been observed, and the exact point along the RISC-assembly pathway at which strand separation occurs is also unknown.[49, 59] In addition, the mechanism by which the guide strand is preferentially loaded onto RISC, whilst the passenger strand is discarded, is also unknown. One theory suggests that Dcr and dsRBPs are able to detect the thermodynamic stability of the binding of each strand to the other and preferentially incorporate one strand over the other. Crosslinking experiments with siRNA modified with 5-iodouracil residues showed that when one siRNA strand has a greater thermodynamic stability at its 5' end than the other, Dcr-2 will bind to the strand with the least stable 5' terminus. R2D2 will bind to the more stable 5' terminus, resulting in the Dcr-bound strand being incorporated into RISC and the R2D2 strand discarded as the passenger strand.[60] This model states that siRNA unwinding occurs prior to RISC incorporation, whereas others have suggested that

strand separation is carried out by Ago2 after the siRNA has been transferred from the Dcr-2/R2D2 complex to RISC.[61]

Following this poorly understood RISC activation stage, the final step of the RNAi pathway is cleavage of the target mRNA strand. This is accomplished by an ATP-independent endonuclease reaction by the Ago proteins, which contain conserved PIWI domains with a structure very similar to RNase H.[62-63] The reaction does not destroy the bound siRNA strand, so RISC is a multiple turnover enzyme, explaining the catalytic nature of RNAi.[49]

Another form of RNAi is driven by microRNAs (miRNA) – hairpin RNA strands that are naturally transcribed by the genome – and is thought to follow a similar mechanism via processing by a Dcr enzyme, and activation of the RISC complex.[49, 59] However, it is unclear how distinct these two pathways are, and how assembly of the micro RISC (miRISC) differs from that of the siRISC. In *D. melanogaster*, separate Dcr enzymes, Dcr-1 and Dcr-2, exist for processing of miRNA and dsRNA respectively,[64] as well as a separate dsRBP, R3D1,[64] and a distinct RISC-associated Ago protein, Ago1.[65] However, in humans only one Dcr enzyme has been indentified, suggesting more integration between the processing of miRNA and siRNA. Sequence complementarity of microRNAs to their target mRNAs is less exact, leading to blocking of translation by mRNA binding rather than silencing by mRNA destruction.[66]

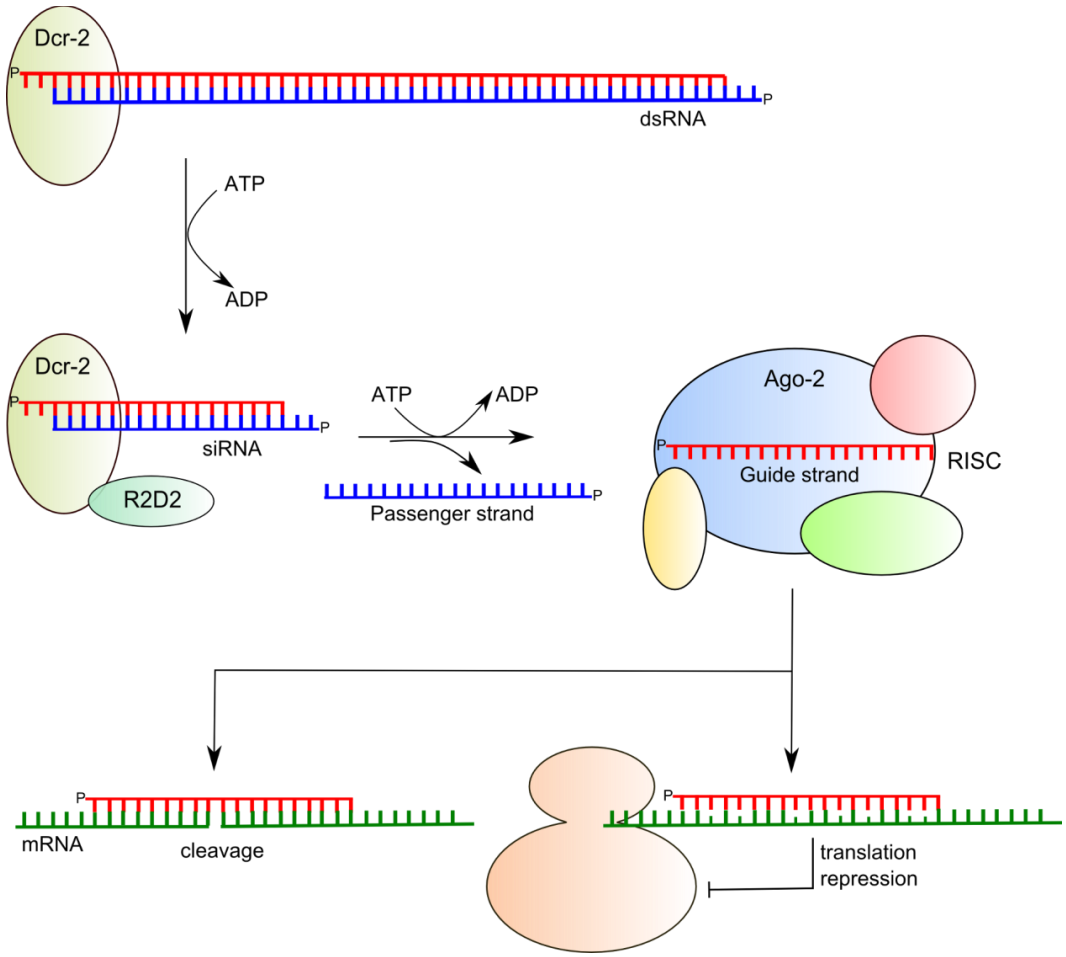


Figure 3. Mechanism of RNA interference in *D. melanogaster*. Double stranded RNA is bound by the Dcr-2 enzyme which cleaves it into ~25 nucleotide siRNA in an ATP-dependent process. The resulting siRNA is loaded onto the RISC complex via a RISC-loading protein called R2D2 which binds both the siRNA and Dcr-2, and the passenger strand is discarded leaving only the guide strand, which is bound by Ago-2. RISC is responsible for loading the guide strand onto its complementary mRNA. Knockdown of mRNA expression occurs either via cleavage of the mRNA by the Ago proteins of RISC if sequence complementarity is high, or via repression of translation once loaded onto the ribosome when complementarity is low, which is usually observed for miRNAs.

1.3.2. Delivery

The discovery of the RNAi mechanism has opened up the possibility of targeting any gene using its complementary siRNA for the treatment of a vast array of diseases. However, before siRNA reaches its target *in vivo*, it faces a number of significant barriers that block its pathway to the RISC machinery. Upon entering the

bloodstream, siRNA is vulnerable to degradation by endogenous nucleases, and renal excretion due to its small size and highly anionic character. In addition, before reaching its target cell, the siRNA must navigate the tight endothelial junctions of the blood vessels and diffuse through the extracellular matrix. Due to its numerous negative charges, siRNA does not readily bind to or cross the cell membrane, and once inside the cells, it must escape from endosomes in order to interact with its intracellular protein targets.[67]

Where siRNA can be applied directly to the target tissue, local delivery of naked, unmodified siRNA has shown to be successful, and several treatments for age-related macular degeneration involving the injection of siRNA directly into the eye have reached clinical trials (Table 1).[8] However, for many therapeutic targets, including metastatic tumours, local delivery is not possible and systemic delivery of siRNA is the only effective route. Therefore, in order to achieve successful delivery of siRNA to these targets, a delivery system is needed which protects the siRNA from degradation by endogenous enzymes and improves the bioavailability of the siRNA to its target cells, enabling the siRNA to cross the cell membrane and escape the endosomes to reach the intracellular machinery.[8, 67-68] The first delivery vectors developed for the delivery of siRNA were viral vectors and were shown to achieve effective mRNA knockdown both *in vivo* and *in vitro*.[69] However, due to safety concerns, such as the observation of mutagenesis caused by insertions,[69] more recent siRNA delivery systems have focussed on non-viral approaches.

1.3.2.1. *Chemical Modifications*

Chemical modifications can be used to increase the efficacy of siRNA delivery *in vivo*. Modifications can be made to oligonucleotides to protect from nuclease degradation (Figure 4) such as addition of a 3' phosphorothioate backbone linkage to protect from exonucleases or 2' modifications such as 2'-*O*-methyl or 2'-fluoro to protect from endonucleases.[8] In addition, the substitution of the 5'-phosphate on the sense strand to a 5'-*O*-methyl prevents the sense strand from binding to RISC, reducing off-target effects.[8]

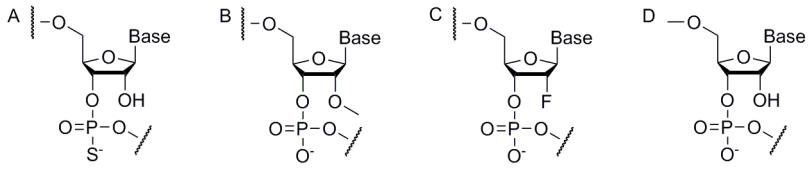


Figure 4. RNA chemical modifications. A: 3'-phosphorothioate to prevent degradation by endonucleases; B and C: 2'-*O*-methyl and 2'-fluoro to prevent degradation by exonucleases; D: 5'-*O*-methyl on the sense strand to prevent binding to RISC.

One type of delivery method involves chemical conjugation of small molecules to the sense strand in order to improve the biodistribution of siRNA. Cholesterol-conjugated siRNA is able to knock down the ApoB gene in mice injected with the conjugate, whereas naked siRNA was found to be ineffective.[70] The cholesterol-siRNA conjugates were also able to bind to human serum albumin, improving biodistribution by restricting clearance by the renal system. The same effect is also observed with various long chain fatty acids and bile acids, which bind albumin as well as high and low density lipoproteins (HDL and LDL). Unconjugated siRNA and siRNA linked to short and medium chain fatty acids do not knock down ApoB *in vivo* and do not bind to albumin or HDL and LDL, which illustrates the importance of lipophilicity in the biodistribution of siRNA.[71] Another type of chemical conjugation has been the attachment of RNA aptamers to siRNA in order to target specific tissues or cell types. For example, an siRNA-RNA aptamer chimera designed to target prostate cancer cells expressing prostate-specific membrane antigen has been shown to knock down cancer survival gene *Plk1* and inhibit tumour growth *in vivo*.[72] Therefore, conjugation of siRNA to small molecules may enhance siRNA delivery by improving bioavailability and targeting, but do not offer any assistance in crossing cell membranes.

1.3.2.2. *Liposomes and Lipoplexes*

One of the most commonly used types of delivery vector for siRNA are liposomes, where the siRNA is encapsulated within a lipid bilayer. These liposomes are formed from synthetic cationic lipids which are composed of a cationic head group, usually a

quaternary ammonium group, and a hydrophobic region consisting of one or two alkyl chains or a steroid such as cholesterol. Delivery of nucleic acids using cationic lipids, known as lipofection, was first demonstrated by Felgner *et al*. using the synthetic lipid *N*-[1-(2,3-dioleoyloxy)propyl]-*N*,*N*,*N*-trimethylammonium chloride (DOTMA).[73-74] This was followed by studies with several other cationic lipids which proved to be successful for *in vitro* delivery of nucleic acids to a wide variety of cell types, including dioctadecylamidoglycylspermine (DOGS),[75] 3β[*N*-(*N'*,*N'*-dimethylaminoethane)-carbamoyl) cholesterol (DC-Chol),[76] dimethyldioctadecyl-ammonium bromide (DDAB),[77] and dimyristyloxypropyl-3-dimethylhydroxyethyl ammonium bromide (DMRIE),[78] as well as commercially available products such as Lipofectamine (Invitrogen), RNAifect (Qiagen) and Oligofectamine (Invitrogen). However, many of the early cationic lipid delivery systems proved to be ineffective *in vivo* and showed high levels of toxicity, with a microarray study of the widely used lipofection agents Lipofectin and Oligofectamine showing alteration of the expression of up to 16% of genes, including genes involved with apoptosis.[79] Another study showed that siRNA complexed with several cationic lipids elicited an inflammatory response in a mice by upregulating the expression of pro-inflammatory cytokines such as IL-2, IFNγ and TNFα.[80]

Cationic lipids are often used in conjunction with other components in order to improve their delivery efficiency. Neutral lipids such as dioleoyl phosphatidyl-ethanolamine (DOPE)[81] or cholesterol[82] are often used as colipids to increase endosomal escape by adopting the inversion hexagonal (H_{II}) phase. Lipids that adopt the H_{II} phase are arranged in hexagonally packed tubes, with each tube containing a narrow aqueous channel in the centre surrounded by the lipid chains on the outside. The H_{II} phase has been shown to be favoured by lipids that participate in membrane fusion due to their decreased stability compared to the bilayer configuration.[83] They are therefore more prone to endosomal escape, making them advantageous for siRNA delivery. However, the incorporation of DOPE into cationic lipid lipoplexes with siRNA leads to instability of the complexes due to salt bridges forming between the phosphate groups of DOPE and the positive charges of the cationic lipid. This

therefore decreases the strength of the interactions between the siRNA and the liposomes and makes the complex more susceptible to destabilisation by acidic serum proteins when used *in vivo*.[84-85]

Many recent studies have focussed on the development of stable lipid delivery systems which show effective gene delivery *in vivo*. In particular, the stable nucleic acid-lipid particles (SNALPs), consisting of small, monodisperse particles with low surface charge have been successful in achieving siRNA-mediated gene silencing in several *in vivo* models. SNALPs were first shown to be effective in targeting the hepatitis B virus (HBV) in mice, where SNALP-formulated siRNA was given in three daily injections of 3 mg/kg/day, resulting in a decrease in HBV levels by 1-2 orders of magnitude.[86] SNALPs have also been shown to be effective in non-human primates, with one study achieving a knockdown of ApoB in monkeys resulting in a 90% decrease in blood ApoB-100 protein two days after a 2.5 mg/kg injection.[87] More recently a study on rhesus monkeys given a lethal dose of Zaire Ebola virus (ZEBOV) showed that the animals could be rescued by seven daily 2 mg/kg treatments of an anti-ZEBOV siRNA cocktail formulated with SNALPs.[88] Other groups are attempting to develop novel cationic peptides to produce more stable lipoplexes by rational design.[89-91] Semple *et al*. have taken a commonly used SNALP cationic lipid, 1,2-dilinoleyloxy-3-dimethylaminopropane (DLinDMA), and used a structure-activity relationship (SAR) approach to optimise the linker and polar head group to produce a new cationic lipid which, when incorporated into SNALPs, achieves a 50-fold increase in ED_{50} in mice and an ED_{50} of 0.3 mg/kg in monkeys.[89] The first RNAi therapeutic using a SNALP has now entered Phase I clinical trials (Table 1), demonstrating the potential of this type of delivery vector.

1.3.2.3. *Cationic Polymers*

Another type of delivery vehicle for siRNA is cationic polymers which form electrostatic complexes, protecting the siRNA from degradation and aiding internalisation by shielding the siRNA negative charges.[68, 92] One of the most widely studied polymers for nucleic acid delivery is polyethyleneimine (PEI, Figure 5a).

PEI is available as both linear and branched isomers, with linear PEI containing only secondary amine groups (except for one terminal primary amine), whereas branched PEI contains primary, secondary and tertiary amines.[92] These ionisable amine groups allow PEI to bind to siRNA, forming stable nanoparticles to protect the siRNA from degradation by serum nucleases.[93] The amine groups also provide PEI with a mechanism for endosomal escape, known as the "proton sponge" hypothesis.[94] The titratable amine groups of PEI are protonated in the endosomes, preventing endosome acidification and causing an influx of chloride ions and water, with the resulting osmotic pressure leading to swelling and lysis of the endosome, releasing the nucleic acid cargo into the cytoplasm. The importance of the ionisable amine groups of PEI was further highlighted by Thomas *et al*. who showed that commercially-available PEI contained ~11% acylated nitrogens as a result of inefficient removal of *N*-propionyl groups from the *N*-propionyl-PEI (PEOZ) precursors.[95] Following complete deacylation of PEI, an increase in its DNA-condensing ability as well as PEI-DNA transfection efficiency was observed. The fully deacylated PEI was used to deliver siRNA targeting the influenza nucleoprotein to mice infected with the influenza virus, resulting in a 94% reduction of virus titre in the lungs after 24 h. Another study used PEI to deliver siRNA targeted to the proto-oncogene c-erbB2/neu (HER-2) both *in vitro*, where a 50% mRNA knockdown was achieved in SKOV-3 cells, and *in vivo*, where significant reduction in tumour growth was observed compared to a non-specific siRNA control using a subcutaneous mouse tumour model.[93] Therefore, PEI shows some potential as a siRNA transfection agent, however its toxicity at higher concentrations resulting from membrane disruption and apoptosis induction may limit its applications as a therapeutic delivery agent.[96]

Other polymers used for siRNA transfection include cyclodextrin-based compounds (Figure 5b). Cyclodextrins are widely used in the food and pharmaceutical industries due to their ability to encapsulate small, hydrophobic molecules, thereby increasing their stability and solubility, making them suitable for drug delivery and food preservation. For use in nucleic acid delivery, cyclodextrins are modified by

addition of cationic polymers such as PEI which complex the oligonucleotides. Cyclodextrin-containing cationic polymers are less cytotoxic and have higher transfection efficiencies than the un-modified polymers.[92, 97]

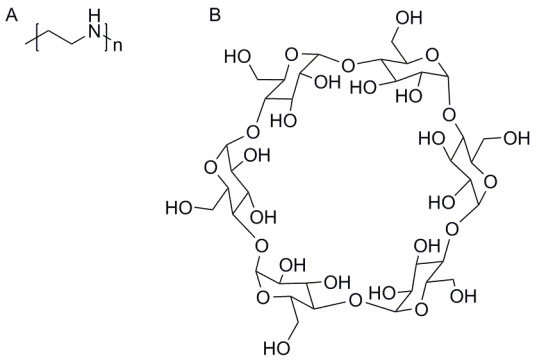


Figure 5. Synthetic polymers for siRNA delivery. A: PEI; B: β-cyclodextrin. This structure is modified by addition of cationic polymers to the hydroxyl groups, allowing complexation of siRNA.

1.3.2.4. *Cell Penetrating Peptides*

CPPs have been employed in the delivery of siRNA both *in vitro* (Table 2) and *in vivo* (Table 3). A few CPPs have been covalently attached to siRNA via disulfide bonds, which are reduced in the intracellular environment allowing interaction of the siRNA with RISC. Successful knockdown was demonstrated by the CPPs penetratin and transportan covalently linked to luciferase or GFP siRNAs.[98] The conjugates were shown to decrease GFP fluorescence in Chinese hamster ovary (CHO) cells expressing GFP by 53% and 63% respectively with knockdown lasting for up to 72 h, compared with Lipofectamine which only achieved a 36% knockdown and lasted for less than 3 days. In EOMA cells expressing GFP, Lipofectamine was also shown to be less effective than the penetratin and transportan conjugates, producing a decrease in fluorescence of 65% compared to 73% and 80% for the CPP-linked siRNA. However, these results should be interpreted with caution, as the continuous CPP-siRNA treatments were compared to Lipofectamine treatments of just 6 hours.[99] In addition, it is not clear whether the knockdown was achieved due to the CPP-siRNA conjugates, or electrostatic complexes since no purification strategy was performed to isolate the disulfide-linked species.

Chiu *et al*. demonstrated that 300 nM of disulfide-linked EGFP siRNA and TAT was capable of delivering up to 70% gene silencing in HeLa cells after a 16 hour incubation, compared to 87% for 20 μg of Lipofectamine.[100] The same CPP was also covalently linked to siRNA for the endogenous gene CDK9, and achieved an 82% knockdown after 42 h at a concentration of 400 nM. This study addressed the issue of whether the knockdown effect was caused by the disulfide conjugate or an electrostatic CPP-siRNA complex by including a control transfection consisting of free siRNA and CPP. This control transfection produced no knockdown, suggesting that the covalently linked species was responsible for the gene silencing. Another study extended the principle of covalently linked CPPs and siRNA to primary cells, using penetratin to deliver siRNA to hippocampal neuron cells.[101] This system achieved knockdown of the endogenous proteins SOD1 and Caspase-3, which were decreased by 90% and 73% respectively at 80 nM according to Western blot analysis. In addition, cell survival using penetratin-siRNA was compared to Lipofectamine, and showed that the penetratin conjugate was much less harmful to neurons, with a 92% survival rate compared to 58% for Lipofectamine. Once again, this study does not purify the CPP-siRNA conjugate, and therefore the species responsible for the observed siRNA delivery is unclear.[99]

More recently, CPP-siRNA conjugates containing penetratin and TAT, have been tested *in vivo* for the first time.[102] An siRNA sequence targeted to p38 MAP kinase, a protein implicated in various inflammatory diseases, was linked to TAT(48-60), penetratin, or cholesterol and the conjugates purified by anion-exchange HPLC and characterised by electrospray mass spectrometry. When incubated with mouse fibroblast cells, these conjugates show a modest 20-36% knockdown of p38 MAP kinase, however when the three conjugates were administered to mouse lungs via intratracheal injections, none of the constructs were able to produce a knockdown significantly different to siRNA alone. In addition, penetratin-siRNA elicited high levels of the immune markers TNFα and IL-12p40, suggesting that this CPP may be responsible for triggering the innate immune system. These results suggest that the

use of covalently-conjugated CPPs might not be ideal delivery vectors for siRNA *in vivo*.

Most of the published studies concerning the delivery of siRNA using CPPs have used the cationic nature of the peptides to form electrostatic complexes, shielding the siRNA negative charges to allow entry into cells. It is thought that these non-covalent complexes may not be as effective as covalent conjugates, as the peptide positive charges also play a role in internalisation by interacting with cell surface proteoglycans (Section 1.2.1.1). Therefore, if these charges are tied up in electrostatic interactions with siRNA molecules, it may not be possible to form sufficient interactions with the cell membrane to allow internalisation.[99] On the other hand, using CPPs in electrostatic complexes may be beneficial as they provide greater protection against nuclease degradation by forming a stable complex. In addition, complex formation is less time consuming as the siRNA requires no modification to allow CPP attachment. Electrostatic complexes remain the most popular application for CPPs in siRNA delivery, and some significant progress has been made both *in vitro* and *in vivo* by several groups using both natural and synthetic CPPs.

One of the earliest such studies examined the siRNA delivery abilities of the amphipathic CPP MPG, a chimera peptide of the hydrophobic fusion peptide domain of HIV-1 gp41 protein and the hydrophilic nuclear localisation sequence (NLS) of SV40 large T antigen.[103] The MPG peptide was successful in internalising siRNA as electrostatic complexes, achieving a 60% GAPDH knockdown at an siRNA concentration of 100 nM after a 30 h transfection in human fibroblast HS-68 cells. However, a mutant MPG peptide possessing a mutation in the NLS region, MPGΔ^{NLS} increased the knockdown to 80% at a siRNA concentration of just 25 nM, supporting the general mechanism that the siRNA targets RISC in the cytoplasm. This early study did not include any cytotoxicity data and also did not compare MPG with any other available transfection reagent.

The use of MPG as an siRNA delivery vector was further explored by Veldhoen *et al*. who formed electrostatic complexes between siRNA and MPGα, a derivative of MPG differing by 6 amino acids in the hydrophobic region increasing its tendency to form a helical conformation.[104] MPGα was compared to Lipofectamine 2000 and showed that whilst the CPP showed a comparable maximum knockdown ability to Lipofectamine, achieving a 70-83% luciferase knockdown in HeLa cells compared to 88-97% for Lipofectamine, its IC_{50} was 30-fold higher than Lipofectamine (0.8 nM compared to 0.02-0.04 nM). By measuring the total internalised siRNA at the half maximal inhibition point of each compound using a highly sensitive liquid hybridisation assay, it was found that around 10,000 molecules were required for half maximal inhibition of MPGα compared to 300 for Lipofectamine. Therefore, the amount of available siRNA was about 30-fold less using MPGα, suggesting that the CPP was either inefficient at escaping from endosomes, or targeted to intracellular compartments other than the cytoplasm, thus preventing the siRNA from binding to RISC. These observations seem to contradict evidence that suggests that MPG-DNA complexes enter cells via a non-endosomal route, and are unaffected by inhibitors of endocytosis.[103] Thus, it appears that very similar peptides may be internalised by different pathways, perhaps influenced by differing sequences, cargoes or particle size

The ability of siRNA complexes to escape the endosomes was studied using a penetratin analogue called EB1 by replacing two amino acids with histidine which, in theory, would promote endosomal escape by the formation of a helical structure upon protonation in the endosome, therefore promoting membrane insertion.[105] The activity of EB1 was compared to the other well characterised CPPs penetratin, $MPG\Delta^{NLS}$, bPrPp, another peptide known to promote endosomal escape,[106] and TP10. In the system tested, EB1, $MPG\Delta^{NLS}$ and bPrPp were able to knockdown luciferase expression to a similar extent to Lipofectamine, achieving a ~50% knockdown with 100 nM siRNA. Under the conditions used, penetratin and TP10 did not show any significant knockdown at 100 nM siRNA. This highlights the importance of localising the siRNA to the cytoplasm for interaction with RISC by

ensuring siRNA can escape from endosomes. Importantly, this study included a toxicity study for all CPPs used, looking at both membrane stability and long term toxicity. Membrane integrity was maintained for siRNA-peptide complexes, but a higher membrane disturbance was observed for uncomplexed peptides, suggesting that the use of electrostatic complexes may have toxicity advantages over covalent conjugates. In a proliferation assay to assess long term toxicity over 48 h, no significant reduction in proliferation was observed for the peptide-treated cells, however Lipofectamine significantly reduced proliferation, showing that CPPs may be more advantageous for *in vivo* siRNA delivery than cationic liposomes.

More recently, in a further attempt to design a CPP with built-in endosomolysis functionality, a report by Andaloussi *et al*. described a modified version of the TP10 CPP called PepFect6, which is functionalised with an N-terminal stearyl group, as well as four covalently attached chloroquine analogues.[107] The N-terminal fatty acid is incorporated to improve serum stability, whilst the chloroquine-like side chains are included to allow the CPP-siRNA complex to escape from endosomes when protonated. The results for PepFect6 show convincing siRNA internalisation across a range of cell lines, including primary cells, which out-perform Lipofectamine with lower toxicity. The peptide also promotes significant knockdown *in vivo* in two mouse models, one using siRNA targeting the ubiquitous endogenous gene hypoxanthine phosphoribosyltransferase 1 (HPRT1) and one in mice expressing luciferase in the liver. The study showed that PepFect6 was capable of achieving knockdown in several organs lasting 72 h after systemic injection of the CPP-siRNA complexes at a concentration of 1 mg/kg. However the most significant knockdown was in the liver, kidneys and lungs, probably due to the high blood supply to these organs, suggesting that delivery to other areas requires more specific targeting (Section 1.3.2.5).

The synthesis of a novel amphipathic peptide further highlighted that not all CPPs are internalised using the same mechanism.[108] CADY, a 20-amino acid peptide containing arginine and tryptophan residues was found to form stable CPP-siRNA

complexes which are capable of >80% mRNA knockdown in 3T3C and HUVEC primary cells at a siRNA concentration of 20 nM and a 40:1 CADY:siRNA ratio. Biophysical analysis of CADY found that the peptide forms an amphipathic α-helix structure upon hydrophobic binding to phospholipid mono- or bilayers, followed by the formation of electrostatic interactions between the polar head groups of the phospholipids with the cationic residues of the peptide. After interaction with the membrane lipids, the helical peptide is then thought to disrupt the membrane, allowing direct penetration of the CPP and cargo across the lipid bilayer and entry into the cell without entrapment in endosomes.[109] It was shown that, like MPG, inhibitors of endocytosis do not reduce CADY-siRNA uptake or reduce mRNA knockdown,[108] suggesting that CADY-siRNA complexes may employ a more direct internalisation mechanism.

One of the simplest cationic peptides, polyarginine, has also been used for siRNA delivery. A 12-mer arginine peptide (R12) was the first homoarginine peptide shown to internalise siRNA and knockdown a target gene in suspension tobacco cells.[110] A recent study using 15-mer arginine (R15) has further characterised the delivery of siRNA both *in vitro* to mammalian cells and *in vivo* to mouse tumours.[111] In Cos-7 cells, R15 achieved a >50% knockdown of luciferase after a 24h incubation with 150 nM siRNA and a R15:siRNA ratio of 3:1. In this case, confocal microscopy suggested that the complexes were taken up via endocytosis into vesicles followed by escape into the cytoplasm. This suggests that polyarginine peptides may have an innate ability to escape endosomes, perhaps due their ability to form interactions with phospholipids and GAGs (Section 1.2.1.1). *In vivo*, 4 μg of siRNA complexed with R15 in a ratio of 3:1 R15:siRNA was administered directly to tumours every 3 days, resulting in significant tumour reduction after 17 days with no observed toxicity. This demonstrates that cationic peptides may have a role in the therapeutic delivery of siRNA if successfully delivered to the cytoplasm.

As mentioned previously, one of the concerns about non-covalent attachment of CPPs to siRNA is that the overall positive charge of the peptides may be neutralised

by the anionic RNA molecules, thus reducing the CPP's ability to bind to negatively charged cell surface GAGs, and therefore the transfection ability of the CPP.[99] Some groups have attempted to modify existing CPPs by attaching RNA-binding components which complex the siRNA leaving the CPP's cationic amino acids available for cell surface interactions. Attaching additional cationic peptides such as polylysine, to CPPs is one technique that has been shown to increase the uptake of several CPPs including TAT, transportan and penetratin.[112-113] It was found that longer attached peptides encourage preferential binding of siRNA to the extended polycationic amino acid chain rather than the original CPP, thus restoring the full transfection capabilities of the CPP. However, significant cytotoxicity was observed when using these fusion peptides, probably as a result of the large positive charges present.

Another CPP modification involves coupling TAT to natural RNA-binding protein domains which are not highly charged. This technique was first described by Endoh *et al*. who linked TAT to the U1 small nuclear ribonucleoprotein A RNA-binding domain (U1A RBD), which when complexed to a modified siRNA reduced fluorescence in CHO cells expressing eGFP by ~65%.[114] However, the approach used in this study had many drawbacks, including the requirement to irradiate the cells with light in order to induce endosomal escape, and the fact that the U1A RBD is a sequence-specific binding domain, and therefore requires a specific RNA sequence to be fused to the siRNA of interest. The concept of using RBDs was further explored using an RNA binding domain from the human dsRNA-activated protein kinase (DRBD), which is a non-sequence specific binding domain, fused to TAT.[115] In this study, significant knockdown of genes in a wide range of cell types was achieved, including in cell lines in which transfection has previously been difficult, such as Jurkat T-cells, HUVEC cells and human embryonic stem (hES) cells. mRNA knockdown by this construct was shown to be significantly greater than with Lipofectamine, and with very low toxicity, and local delivery *in vivo* was also successful. This study shows that the use of RNA binding domains to complex RNA, may be a useful technique for improving uptake of currently used CPPs,

perhaps by freeing up the peptides' basic amino acids to allow interaction with anionic cell membrane GAGs. However, this technique depends upon the expression and purification of recombinant proteins, which may be time consuming and expensive.

Table 2. CPPs used for *in vitro* siRNA delivery.

CPP	Sequence	Attachment	Cell Lines	Target Genes	siRNA Conc (IC_{50})	Cytotox. data available	Refs
Penetratin	CRQIKIWFQNRRMKWKK	Covalent	C166, EOMA, CHO Hippocampal Neurons	eGFP, Luciferase SOD1, Caspase3	25 nM 80 nM	No Yes	[98] [101]
Transportan	CLIKKALAALAKLNIKLLYGASNLTWG	Covalent	C166, EOMA, CHO	eGFP, Luciferase	25 nM	No	[98]
TAT_{47-57}	CYGRKKRRERRR	Covalent	HeLa	eGFP, CDK9	300 nM	No	[100]
MPG **$MPG\Delta^{NLS}$** **MPG-8**	GALFLGFLGAAGSTMGAWSQPKKKRKV GALFLGFLGAAGSTMGAWSQPKSKRKV AFLGWLGAWGTMGWSPKKKRK	Non-covalent	HS-68, HeLa, Cos-7 HS-68, HeLa, MCF-7, PC3, SKBr3-HER2	Luciferase, GAPDH Cyclin-B1	25-100 nM 0.125-20 nM (0.6-1.2 nM)	No Yes	[103] [116]
MPGα	GALFLAFLAAALSLMGLWSQPKKKRKV	Non-covalent	HeLa, ECV304	Luciferase	50 nM (0.8 nM)	No	[104]
EB1	LIRLWSHLIHIWFQNRRLKWKKK	Non-covalent	HeLa, GepG2	Luciferase	100 nM	Yes	[105]
CADY	GLWRALWRLLRSLWRLLWRA	Non-covalent	U_2OS, THP1, 3T3C, HUVEC	GAPDH	0.5-40 nM (1.2-3.4 nM)	Yes	[108]
Polyarginine	R15 Cholesterol-R9	Non-covalent	Cos-7 CT-26	Luciferase VEGF	150-300 nM -	No Yes	[111] [117]
TAT-RBD	TAT-U1A TAT-DRBD	Non-covalent	CHO H1299, Jurkat T-cells, murine T-cells, HUVEC, hES	eGFP dGFP, GAPDH, CD4, CD8, OCT4	200nM 100-400 nM	Yes Yes	[114] [115]
PepFect6	St-AGYLLGK(ε-K_3qn_4)INLKALAALAKKIL	Non-covalent	HEK293, U_2OS, HepG2, CHO, Hepa1c1c7, MEF, Bhk21, N2a, SHSY5Y, B16, U87, RD4, K562, C17.2, HUVEC, Jurkat, mES	Luciferase, eGFP, HPRT1, OCT4	3-100nM (2.6-29.3 nM)	Yes	[107]

Table 3. CPPs used for *in vivo* siRNA delivery

CPP	Attachment	Model	Target Gene	Dose (siRNA)	Delivery Route	Refs.
R15	Non-covalent	Ovarian tumour xenograft	HER-2	4 μg /mouse	Intratumoral	[111]
Cholesterol-R9	Non-covalent	Colon adenocarcinoma xenograft	VEGF	3.5 μg /mouse	Intratumoral	[117]
PepFect6	Non-covalent	Mice, Luciferase transgenic mice	HPRT1, Luciferase	1 mg/kg	Intravenous	[107]
TAT-DRBD	Non-covalent	Luciferase transgenic mice	Luciferase	~10 μg /mouse	Intranasal	[115]
Cholesterol-MPG-8	Non-covalent	Prostate/lung carcinoma xenograft	Cyclin-B1	5-10 μg /mouse	Intratumoral/Intravenous	[116]

1.3.2.5. *Nanocarriers*

Although many of the siRNA delivery strategies described above have been successful in delivering functional siRNA both *in vitro* and sometimes *in vivo*, very few have progressed to use in clinical trials. In fact, most of the first siRNA-based therapeutics to enter clinical trials involved local administration of naked siRNA to the target site, with more advanced delivery systems only appearing more recently (Table 1). Thus, efficient siRNA delivery still remains a difficult challenge, with the goal of finding a vector that provides effective stabilisation of siRNA, cell internalisation, endosomal escape and low toxicity yet to be reached. Rather than using just one of the aforementioned delivery systems, researchers are increasingly using rational design strategies to build complex, multi-component nanoparticles to achieve siRNA delivery for therapeutic applications.[8]

One classification system for these siRNA nanoparticles was described by Kostarelos and Miller as the ABCD system, where a nucleic acid (A) is complexed with layers of polymeric material which aids cell entry (B), biological stabilising compounds (C) and cell targeting ligands (D) (Figure 6).[118] There are many examples of nanocarriers that are comprised of some or all of these components, and there are several thorough reviews that summarise recent advances in siRNA delivery technology.[8, 119-120] In general, the B component is cationic, such as cationic liposomes, synthetic polymers or cell penetrating peptides, as described previously. This layer binds to siRNA electrostatically and promotes cell entry and endosomal escape. Layer C is used to confer additional serum stability to the nanocarrier and increase its half life *in vivo*. PEG is often used as a biological stabiliser since it neutralises the positive surface charge of the cationic liposome or polymer, resulting in decreased interactions with serum proteins, which result in inactivation and aggregation of the siRNA nanoparticles. A targeting component, D, is often included in nanoparticles intended for systemic delivery. This component can be used to specifically target certain cell types, for example using antibodies, or ligands with specificity for particular cell-surface receptors, such as RGD peptides which are used to target cancer cells overexpressing cell surface integrin proteins.

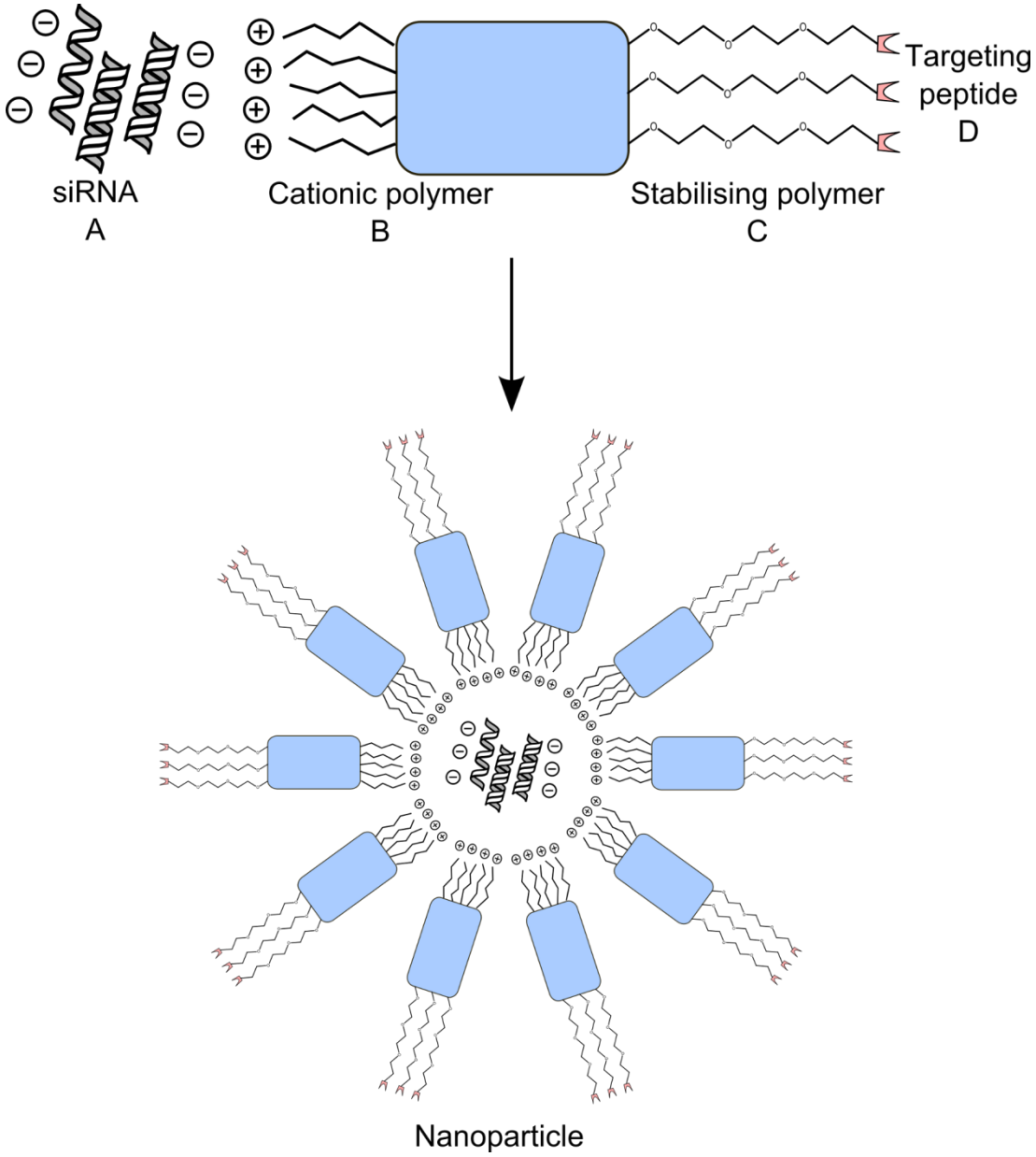


Figure 6. Structure of ABCD siRNA nanoparticles. A is the desired siRNA for a particular target gene; B is a cationic polymer used to shield the negative charges on the siRNA and to promote uptake into cells; C is a stabilising polymer such as PEG to increase stability of the complex to degradation by serum proteins; D is a targeting ligand used to confer specificity for certain cell types.

One example of a nanocarrier system was developed by Rozema *et al*. who named their delivery vehicles siRNA Dynamic PolyConjugates.[121] This system was based around an amphipathic polymer (B) called PVABE, linked to PEG as a biological

stabiliser (C) and a hepatocyte targeting ligand (D), with the siRNA covalently conjugated to the polymer by a disulfide bond. The targeting and PEG moieties are attached to PVABE via maleamate linkages to the amine groups of the polymer, forming negatively charged nanoparticles which do not bind to serum proteins. Following internalisation by endocytosis, the maleamate bonds are readily hydrolysed on acidification of the endosome, exposing the cationic amine groups of the polymer and inducing endosomal escape via the proton sponge effect. When PEG was attached via a non-hydrolysable amide linkage, knockdown ability of the polyconjugates was abolished, showing that exposure of the cationic residues of the polymer are necessary for successful transfection. The polyconjugates were tested both *in vitro*, knocking down apoB in mouse hepatocytes, and *in vivo*, targeting apoB and ppara-1 in mouse livers. In cell cultures, an 80% mRNA knockdown was achieved using ~50 nM siRNA conjugated to polyconjugates, which is comparable to the CPPs listed in Table 2. *In vivo*, an 80-90% knockdown of apoB was achieved using a 2.5 mg/kg dose, again comparable to both CPPs and SNALPs, with very little toxicity. It was found that the presence of the hepatocyte targeting ligand *N*-acetylgalactosamine (NAG) resulted in higher uptake in hepatocytes, whilst replacement of NAG with mannose resulted in targeting of nonparenchymal cells in the liver, confirming the ability of targeting ligands to direct the polyconjugates to the required cell type.

Effective delivery of siRNA remains one of the great challenges currently facing biological therapeutics. The exploitation of the RNAi pathway holds huge potential for the treatment of a wide array of diseases, including cancers, viral infections and genetic disorders. The key components of a successful siRNA delivery agent are the ability to stabilise siRNA in the body, target it to the tissue of interest and deliver it efficiently to the cytoplasm with minimal side effects and toxicity. A huge number of potential delivery vehicles for siRNA have been described in the literature, and the research interest in this area continues to grow year on year. As we discover more efficient ways to achieve siRNA transfection, the prospect of a new era of genetic medicines draws closer.

1.4. SMoCs

Small molecule carriers (SMoCs) are a class of amphipathic, α-helix mimics based on penetratin, consisting of a hydrophobic biphenyl backbone linked to hydrophilic guanidine side chains.[122] The biphenyl system has been found by molecular modelling studies to mimic an alpha helix structure, with positions 2 and 2' on the biphenyl system being spatially equivalent to the *i* and *i+3* or *i+4* helix positions (Figure 7).[123]

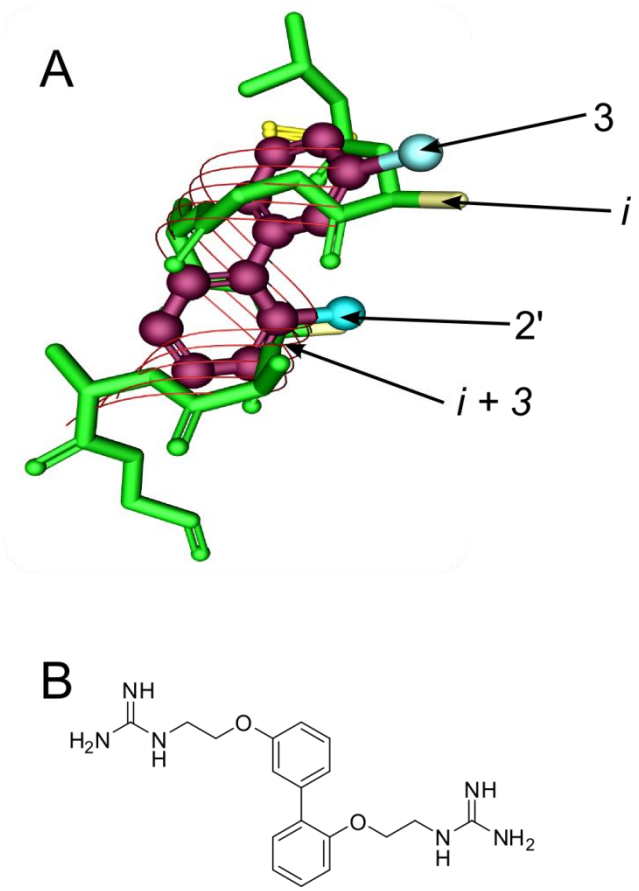


Figure 7. SMoC structure based around a biphenyl ring system. A: Overlay of a biphenyl structure and an α-helix showing that the 3 and 2' positions on the biphenyl are spatially equivalent to the *i* and *i + 3* positions on the helix. (Image reproduced with permission from ref. [122]). B: A 2G-SMoC compound with guanidine-displaying side chains positioned at the 3 and 2' positions on the phenyl ring. This compound should therefore be equivalent to a helical peptide containing two arginine residues.

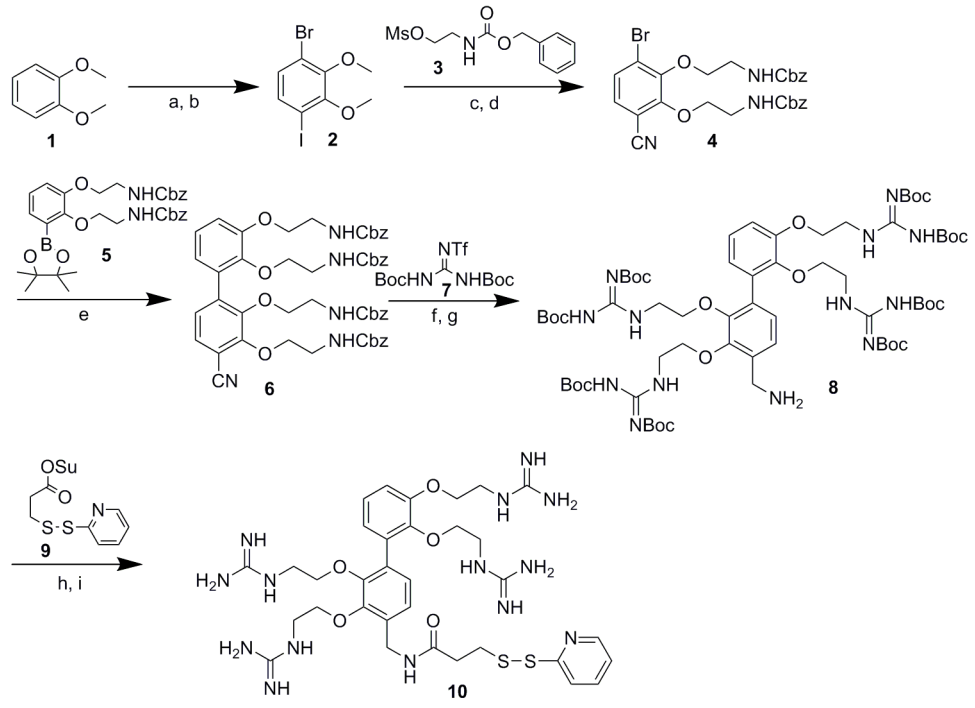
Guanidine groups attached to the biphenyl structure mimic the arginine side chains displayed on the surface of natural CPPs which are important for cell-surface binding. SMoCs also contain a linker group consisting of a free amine which can be used to couple dyes for visualisation by confocal microscopy. In addition, a thiopyridyl derivative has been produced which may be used for covalent conjugation of proteins via a disulfide linkage to a cysteine residue.

Various analogues of SMoC have been produced, with varying numbers and positions of guanidine groups, and varying cargo linker positions.[124] The SMoC analogues are named according to the number of guanidine (G) groups attached to the biphenyl backbone, for example 4G-SMoC denotes a SMoC compound with four guanidines. Increasing the number of guanidine groups has been found to increase the efficiency of transport across cell membranes, with 4G-SMoC possessing the highest activity, possibly due to increased interactions with negatively charged membrane GAGs. The position of the linker is also important, with steric effects decreasing transport efficiency when placed *para* to the biphenyl bond, rather than *meta*, where the macromolecular cargo is furthest from the guanidine groups. This further demonstrates the significance of the interactions of the guanidine groups with the cell membrane.

1.4.1. Synthesis

The current synthetic route to SMoC compounds (Scheme 1) centres around the coupling of the respective mono-phenyl components substituted at the required positions with Cbz-protected amines via Suzuki-Miyaura Pd(0) coupling chemistry. A key intermediate is the iodobromo-derivative **2** of veratrole **1**, which was synthesized by sequential *ortho* metallations. The methoxy groups are cleaved with boron tribromide, and a bis-alkylation is performed with 2-(benzyloxycarbonylamino)ethyl methanesulfonate **3** in order to introduce two Cbz-protected amine groups. The palladium-catalysed cyanation of the iodine with $Zn(CN)_2$ is carried out to produce the nitrile compound **4**. This intermediate is then coupled to the boronic ester **5**, also substituted with two Cbz-protected amines,

which is synthesized from bromoveratrole, to produce the Suzuki product **6**. The amines are then deprotected with hydrogen bromide following coupling, and Boc-protected guanidine groups are added using the guanidinylation reagent *N*,*N*-di-Boc-*N'*-trifluoromethanesulfonyl-guanidine. The nitrile is reduced using Raney-Nickel under hydrogen and in the presence of ammonia to produce the free amine **8** which may be used as a linker for the coupling of dyes, or *N*-succinimidyl-3-(2-pyridyldithio)propionate (SPDP) **9** to add the thiopyridyl (SSPy) protein linker. Finally the guanidines are deprotected using trifluoroacetic acid (TFA) to yield the final 4G-SMoC-SSPy **10**.



Scheme 1. Current synthetic route to 4G-SMoC-SSPy. Reagents: a) BuLi, TMEDA, Et_2O, 1h, room temperature, then, TMSCl, 16h, room temperature; then BuLi, TMEDA, 2h, room temperature; then $C_2Br_2Cl_4$, 1h, -40°C; then warm to room temperature (89%); b) ICl, DCM, 2h, -40°C (26%); c) BBr_3, DCM, 16h, room temperature; then **3**, DMF, Na_2CO_3, 24h, 80°C (59%); d) $Zn(CN)_2$, Pd_2dba_3, dppf, $Zn(OAc)_2$, Zn, DMA, 140°C, 10min, MW (61%); e) 5, $PdCl_2dppf$, K_3PO_4, toluene, water, 18h, 100°C (100%); f) HBr, DCM, 90min, room temperature; then **7**, Et_3N, DCM, 16h, room temperature (65%); g) Raney-Ni, NH_4OH, THF, H_2, 16h (44%); h) **9**, DIEA, DCM, 16h, room temperature (11%); i) TFA/H_2O/TIPS (95:2.5:2.5), 3h, room temperature (100%).

1.4.2. Translocation activity

SMoCs have been shown to internalise both dyes and proteins into a variety of cell types.[122] 2G- and 4G- SMoCs were coupled to fluorescein, and were found to be internalised into U2OS cells and were observed in cytoplasmic vesicles, suggesting an endocytotic mechanism. 4G-SMoC coupled to Alexa Fluor 488-tagged geminin was also successfully internalised as measured by fluorescence-activated cell sorting (FACS) analysis. When cells were treated with chlorpromazine, an inhibitor of clathrin-mediated endocytosis, uptake of 4G-SMoC-geminin-Alexa Fluor 488 was reduced by 70% suggesting an endocytotic mechanism for SMoC-coupled proteins. Geminin is a cell cycle regulator, binding to Cdt1 to prevent loading onto the MCM complex, thus blocking licensing of the DNA replication origin.[125] Untagged geminin was linked to 4G-SMoC via a disulfide linkage and its ability to suppress DNA replication was analysed by measuring BrdU incorporation. Successful delivery of active geminin was observed in a variety of cell types including NIH-3T3 fibroblasts, WI-38 HDF cells, HeLa cells and MOLT-4 leukemic lymphoblasts.[122]

1.5. Thesis Overview

Biological therapeutics has become an area of intense research interest over the last two decades, and the discovery of the RNAi pathway has opened up the possibility of effective gene therapy for a wide range of dieseases. The development of a cellular transporter is one of the major challenges to the progress of macromolecular therapeutics, and an efficient, safe method of delivering proteins and nucleic acids to the intracellular environment would open up many possibilities for new drugs and treatment strategies. While progress has been made in the search for delivery vectors for macromolecular cargoes, significant challenges still remain. Large polymeric carriers such as CPPs have a high cost of manufacture and are also susceptible to degradation by serum proteases, requiring PEG modification in order to achieve sufficient bioavailability. Cationic liposomes and cationic polymers such as PEI are unsuitable for *in vivo* usage, due to their high toxicity. A small molecule transporter such as SMoCs would be desirable due to their low cost of production and serum stability.

The synthesis of the SMoC compounds has been optimised since the first published method,[124] however some steps are still low yielding, in particular the formation of the boronic ester **5** for the Suzuki coupling step which has a yield of 30%, or expensive, such as the guanidinylation step using *N*,*N*-di-Boc-*N'*-trifluoromethane-sulfonylguanidine. It is hypothesized that more effective reagents exist which will increase the yield of these steps, such as Molander-type reagents for Suzuki coupling,[126] and therefore the first aim of this thesis is to find more effective Suzuki and guanidinylation reagents in order to produce higher quanitities of SMoCs. In addition, it was shown that SMoCs with the linker in the *meta* position to the biphenyl bond were more efficient protein transporters than those with the linker at the *para* position, however a *meta* derivative of 4G-SMoC was not produced. Since 4G-SMoC was found to be the most efficient compound for cell internalisation, it would be interesting to synthesize an analogue with a *meta* linker to determine whether uptake could be increased.

Based on studies of CPPs, it is likely that the SMoC compounds bind to the GAGs of cell surface proteoglycans as a precursor to internalisation. Guanidine groups are able to form both bidentate hydrogen bonds and electrostatic interactions to GAGs and phospholipids. In addition, guanidinium cations may also be used to form electrostatic complexes with siRNA molecules. It is hypothesized that the biphenyl SMoC backbone may give SMoCs an advantage over CPPs by stabilising the guanidine positive charges via interactions between the π-electron cloud of the biphenyl ring system and the cationic guanidine groups, so-called π-cation interactions. The second aim of this thesis is to further characterise the SMoC compounds in terms of their internal structure in order to understand how SMoCs bind to anionic molecules such as GAGs and siRNA. The interactions between SMoCs and siRNA will be studied in detail in order to determine the most efficient SMoC compound for complexing siRNA.

Previous studies have shown that SMoCs function as molecular transporters, crossing the cell membrane and delivering proteins which are able to exert a biological effect. Due to their similarity to the CPP penetratin, it is hypothesized that SMoCs may be used to transport additional macromolecular cargoes such as siRNA into cells in order to achieve gene-specific knockdown. The third aim of this thesis is to use SMoC-siRNA complexes to demonstrate that SMoCs may act as siRNA delivery agents *in vitro* by performing transfection experiments in cultured cells.

Finally, it is hypothesized that GAG clustering by SMoCs may lead to internalisation via macropinocytosis as suggested for CPPs, and therefore by increasing the ability of SMoCs to bind and cluster common GAG chains, a greater level of cellular uptake may be achieved. Therefore the final aim of this thesis is to design and synthesize a new SMoC compound in order to more effectively bind cell surface proteoglycans, and to compare the siRNA delivery properties of this SMoC against previous compounds.

CHAPTER 2

Improving the synthetic route to the SMoCs

2. Improving the synthetic route to the SMoCs

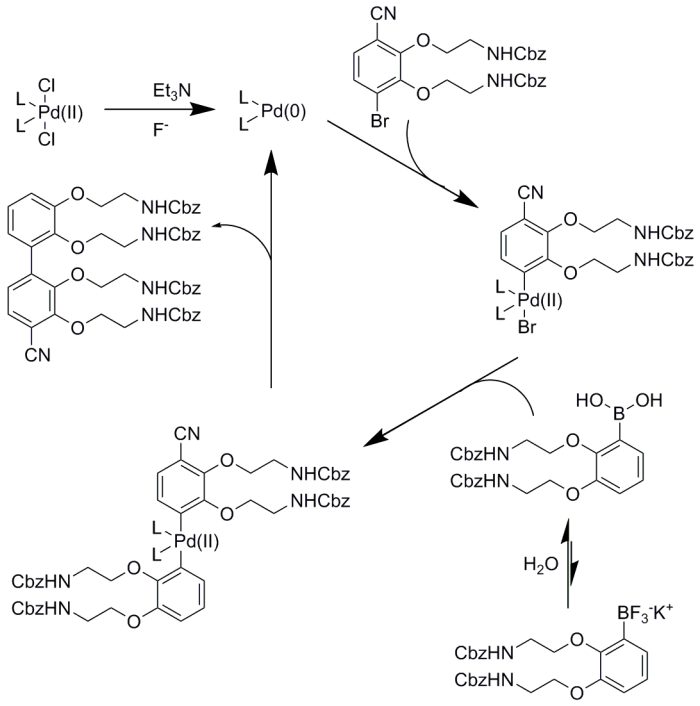
2.1. Introduction

The previously described method for synthesizing 4G-SMoC (Scheme 1) has so far only yielded small quantities of the final product, with some steps producing unsatisfactory yields, such as the synthesis of the boronic ester for the Suzuki-Miyaura coupling which gives a 30% conversion, and the nitrile reduction step, yielding 44% of product.[124] The guanidinylation step uses an expensive triflate reagent which also limits the total amount of SMoC that can be produced due to the cost of materials. In order to produce larger quantities of SMoC to enable more extensive biological testing, these problematic steps should be optimised. In addition, in order to increase the speed of production, stable intermediates should be produced on a large scale which can be easily isolated and stored for future use.

In order to increase the yield of the guanidinylation step, two routes will be explored. Firstly, it is hypothesized that it may be possible to reduce the overall number of steps in the SMoC synthetic route by producing an alkylation reagent bearing a protected guanidine group. This reagent will use the pentamethylchromansulfonyl (PMC) protecting group which is commonly used in peptide synthesis,[127] and may be used to alkylate the catechol precusor of 4G-SMoC. This will eliminate the Cbz deprotection and subsequent gunaidinylation steps. An alternative method for increasing the guanidinylation yield is to use an alternative reagent to the triflate reagent **7**, which if successful will decrease the cost of SMoC production.

The low yielding synthesis of the boronic ester may be overcome by using an alternative reagent for the Suzuki coupling reaction. One such example is the equivalent boronic acid, which is more reactive and therefore may also increase the yield of the coupling. In addition, it is possible to use the trifluoroborate salt. Aryltrifluoroborate salts, developed by Molander *et al*.,[126, 128] are extremely stable to air and moisture, allowing large scale synthesis of this key intermediate for long-term storage. These salts are converted to the boronic acids *in situ* during the reaction, and it is this species that is coupled to the halide, due to the lower energy barrier

involved in the transmetallation step for the boronic acid.[129] The trifluoroborate salts are more efficient than boronic esters in Suzuki reactions due to the significantly lower production of side products formed by homocoupling and oxidation reactions.[129] Homocoupling side products are often formed by Suzuki reactions where a palladium(II) precatalyst is used, which is then reduced to Pd(0) in situ using a base. Before it has been reduced, the aryl boronic acid substrate may activate the catalyst, resulting in two transmetallation exchanges followed by reductive elimination of the homocoupled product. When a trifluoroborate salt is used, fluoride catalyses the reduction of the precatalyst before the boronic acid has formed, for example by forming a fluorophosphonium species which is rapidly hydrolysed.[130]



Scheme 2. Mechanism of 4G-SMoC Suzuki-Miyaura coupling using potassium trifluoroborate salt. The fluoride ions in the solution from the salt catalyse the reduction of the Pd(II) precatalyst to the active Pd(0) form, preventing the homocoupling reaction which occurs via transmetallation onto the precatalyst.[129] The halide then adds to the catalyst, followed by transmetallation of aryl group from the boron to the palladium. Reductive elimination of the two ligands gives the coupled product and reforms the Pd(0) catalyst.

Therefore, using the trifluoroborate salt, it is hypothesized that the Suzuki-Miyaura coupling of the SMoC intermediates should produce fewer side products, resulting in a higher yield (Scheme 2).

As shown in the previous publication on SMoCs,[124] the use of a linker positioned *meta* to the biphenyl bond results in more efficient cellular uptake due to decreased steric interference between the linker and the guanidine groups, however a 4G-SMoC derivative with such a linker was not synthesized. In addition, the attachment of the thiopyridyl linker for attachment to proteins via a disulfide bond is also a low yielding step at only 11%. It is hypothezed that the *meta*-linked 4G-SMoC may be more effective at internalising cargo into cells. Therefore, the sythesis of this compound will be explored, as well as the possibilities of alternative linker chemistries to improve the yield of the final product.

Finally, it was found that upon storage, 4G-SMoC-SSPy attached to the thiopyridyl protein linker dimerises via disulfide bonds, producing a mixture of monomer and dimer SMoCs. It is thought that one of these compounds may be the active species, and therefore, the dimer should be synthesized and purified.

The aims of this chapter are therefore to increase the yield of the SMoC guanidinylation step; to synthesize alternative Suzuki coupling reagents; to synthesize a 4G-SMoC derivative with a linker group *meta* to the biphenyl bond; and to synthesize and purify a disulfide-linked 4G-SMoC dimer.

2.2. Results

2.2.1. Modifications to the synthetic route to 4G-SMoC

2.2.1.1. *Use of PMC protecting group*

The PMC group was used to protect a guanidine group linked to a pyrazole group using the guanidinylation reagent 1*H*-pyrazole-1-carboxamidine in the presence of an organic base.[131] This protected guanidine **12**, which was formed in 43% yield, was

then used to guanidinylate the amine of ethanolamine in an S_N2 reaction with the pyrazole as the leaving group, forming the basis of a PMC-protected SMoC side chain **13** in 40% yield (Scheme 3).[131] The new guanidinylation reagent was characterised by NMR (Figure 8).

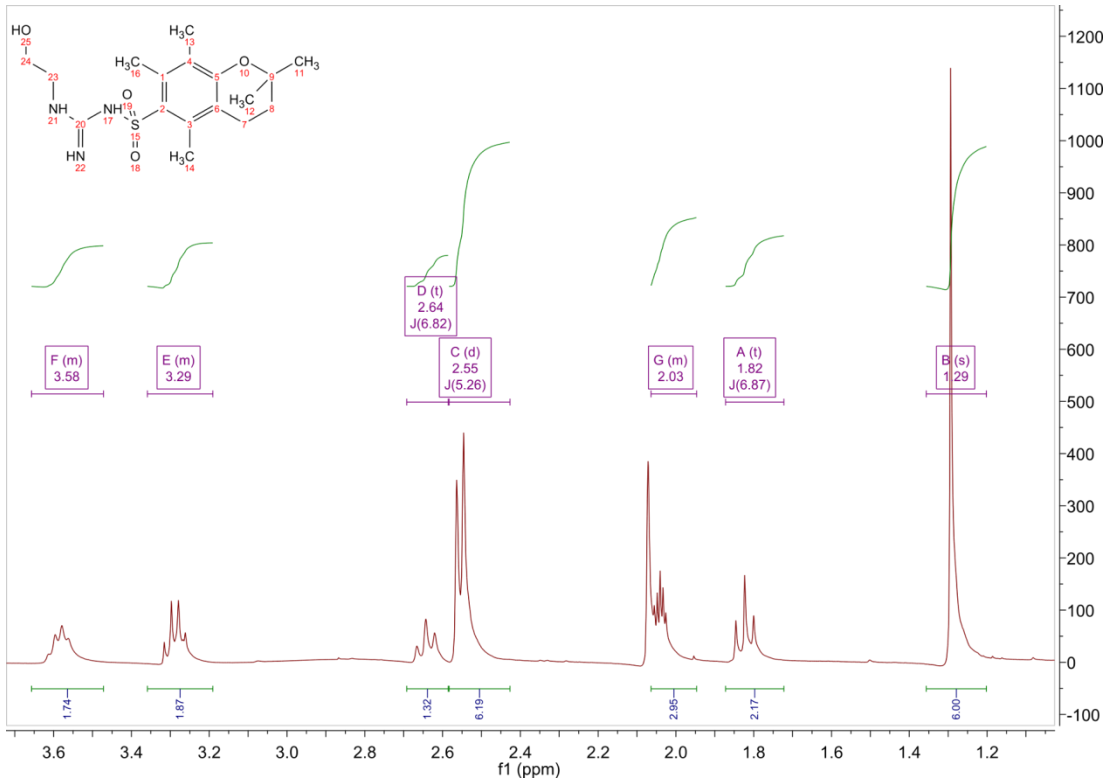
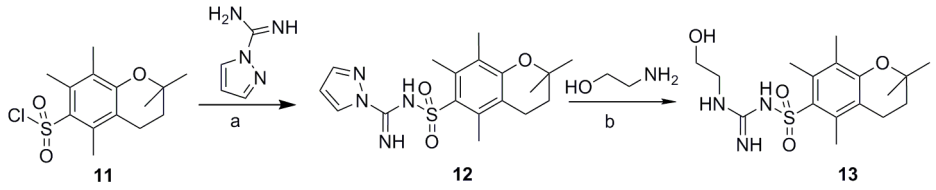
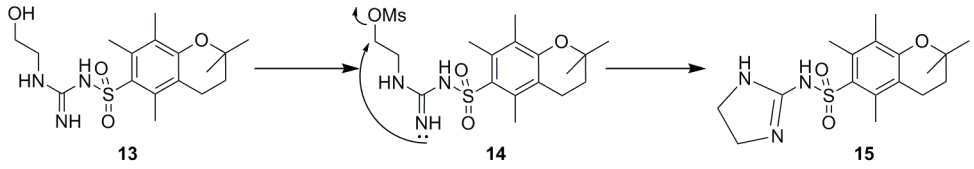


Figure 8. 300 MHz ^{1}H-NMR spectrum PMC-protected guanidinylation reagent 13. The singlet at 1.29 ppm corresponds to the two methyl groups joined to the pyran ring; the triplet at 1.82 ppm which is coupled to the triplet at 2.64 ppm with a *J* coupling constant of 6.8-6.9 Hz correspond to the pyran methylene groups; the multiplet at 2.03 ppm corresponds to the methyl group at position 1 of the aromatic ring, and the two singlets at 2.55 ppm correspond to the other two methyl groups on the aromatic ring; finally, the two deshielded multiplets correspond to the two methylene groups next to the guanidine (3.29 ppm) and the alcohol (3.58 ppm) respectively.



Scheme 3. Synthesis of a PMC-protected guanidine derivative. Reagents: a) 1*H*-pyrazole-1-carboxamidine, DIEA, dioxane, 50°C, 17h (43%); b) Ethanolamine, dioxane, rt, 24h (40%).

In order for the alcohol **13** to be attached to the SMoC ring structure, it was necessary to convert it to the mesylate compound in order to participate in the alkylation of the catechol derived from **2**. Conversion to the corresponding mesylate under standard mesylation conditions of stirring the alcohol in DCM with triethylamine with mesyl chloride[132] failed due to the rapid cyclisation of the mesylate in the presence of base (Scheme 4) due to attack of the free imine on the carbon adjacent to the mesylate forming the imidazoline **15**. This was formed preferentially over the mesylate with a product ratio of 2:1 (**15**:**14**) as detected by LC-MS (Figure 9).



Scheme 4. Cyclisation of alcohol 13 following mesylation. See Table 4 for reagents.

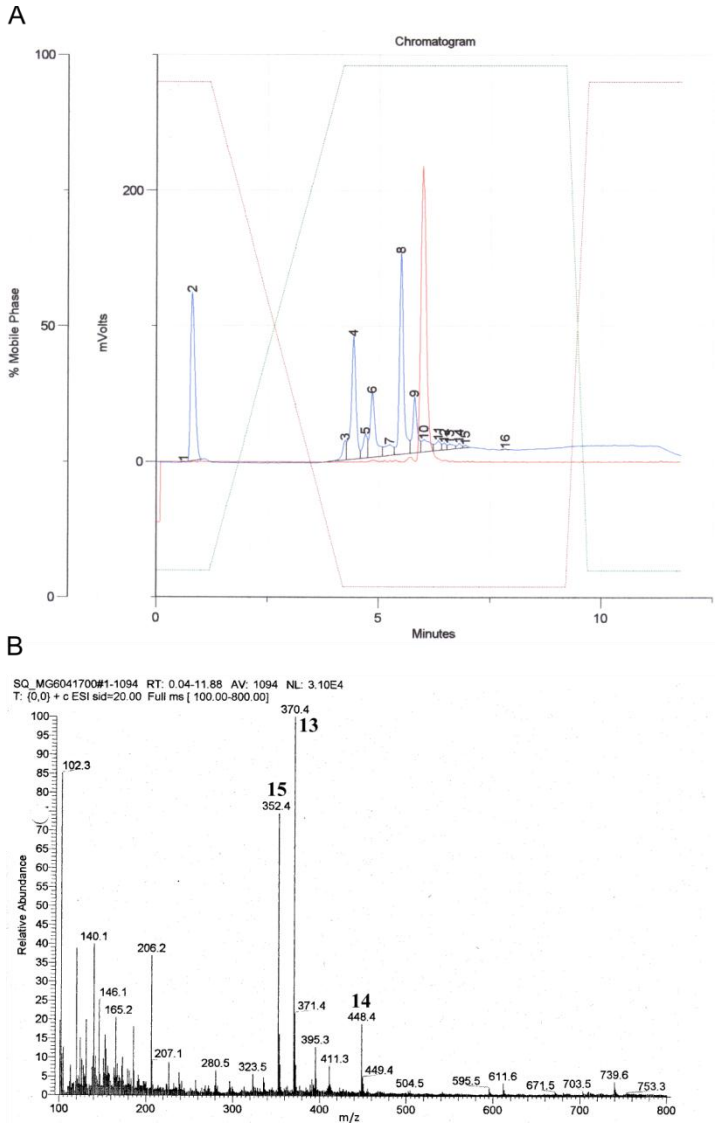


Figure 9. LC-MS analysis of the mesylation of alcohol 13 reaction mixture. A: UV trace of the reaction mixture (blue line), peak 4 is the cyclised product **15**, peak 6 is the mesylate **14**, and peak 8 is the starting material **13**; B: Mass spectrum of the reaction mixture showing the $[M + H]^+$ ions of **13** (*m/z* 370.4), **14** (*m/z* 448.4) and **15** (*m/z* 352.4).

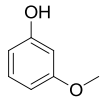
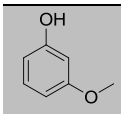
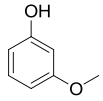
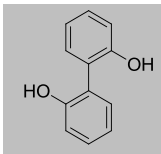
The mesylation reaction was attempted under several different conditions, including using two equivalents of MsCl in order to drive the mesylation reaction to completion before the cyclisation could occur, and using a weaker base (Na_2CO_3), in order to prevent the cyclisation (Table 4). None of the conditions were successful, and so this line of enquiry was abandoned.

Table 4. Conditions tried for the mesylation of 13

Base	Temp. °C	Eq. MsCl	Solvent
Et_3N	Rt	1	DCM
Et_3N	Rt	2	DCM
Na_2CO_3	Rt	1	DCM

As an alternative to alkylation using the mesylate, it was decided to investigate the use of Mitsunobu chemistry which allows alcohols to be coupled to nucleophiles, to attach the alcohol, **13**, to phenol derivatives. A series of test reactions were carried out using 3-methoxyphenol as the nucleophilic species (Table 5). The first reaction used the standard Mitsunobu reagent diisopropylazodicarboxyl-ate (DIAD) and triphenylphosphine (TPP) in THF, which is a widely reported method for this reaction,[133] however no reaction was observed. The more forcing reagents of *N*,*N*,*N'*,*N'*-tetraisopropylazodicarboxamide (TMAD) and tributylphosphine, which are often used for less acidic nucleophiles with higher pK_a values due to the increased basicity of the azo nitrogen atoms,[133] were then tried, using one, and then two equivalents of the reagents. When still no reaction was seen, a biphenyl starting material was tested with the same reagents, with the possibility that the guanidines could be added following the Suzuki-Miyaura coupling step. In this reaction, a small amount of the desired product was observed in the reaction mixture, but the low conversion was not sufficient for further progress.

Table 5. Reagents and conditions attempted for the coupling of 13 to phenols

Phenol	Coupling reagents	Solvent	Conditions	Result
	DIAD, PPh_3	THF	Rt, 24 h	No reaction
	TMAD, PBu_3	THF	Reflux, 48 h, N_2	No reaction
	TMAD, PBu_3 (2 eq each)	THF	Rt, 24 h	No reaction
	TMAD, PBu_3 (2 eq each)	THF	Rt, 24 h	Small LC/MS peak for mono-product

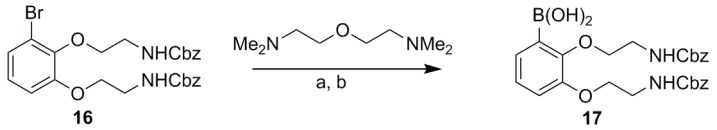
Although several other Mitsunobu reagents are available,[133] the fact that no reaction was observed for the harsh TMAD regent, as well as time constraints meant that this line of enquiry was abandoned in favour of the existing method of using Cbz-protected amines.

2.2.1.2. *Improving Suzuki-Miyaura coupling*

NB. The work in this section was on going in the lab at the same time as the other experiments described in this chapter and was carried out by Roberta Worthington. It is included here for context as part of the overall development of the SMoC synthesis, and will be used later in developing new SMoC reagents.

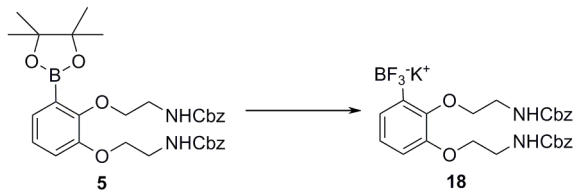
As an alternative to the boronic ester, the equivalent boronic acid **17** was synthesized from the aryl bromide using a Grignard reagent with an organic ligand followed by quenching with trimethyl borate, forming the boronic acid in 40% yield (Scheme 5). The organic ligand moderates the reactivity of the arylmagnesium chloride in this

reaction, reducing possible side reactions.[134] Careful control of the reaction temperature was required to obtain a synthetically useful yield.



Scheme 5. Boronic acid synthesis. Reagents: a) Bis[2-(*N*,*N*-dimethylamino)ethyl] ether, ^{i}PrMgCl, THF, 10-15°C, 24 h; b) Trimethyl borate, THF, rt, 1 h (40%). Reaction carried out by Roberta Worthington.

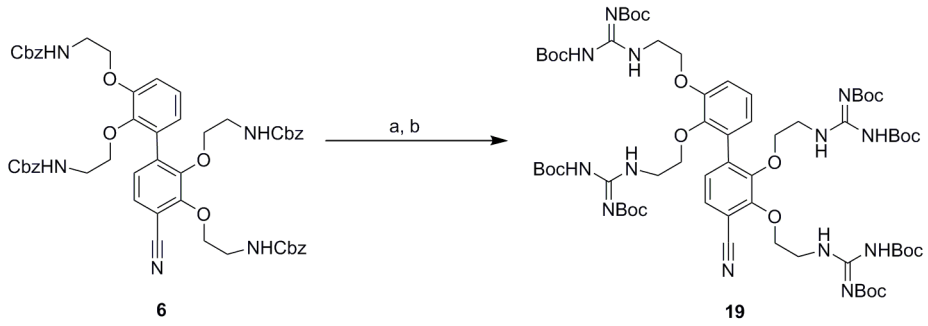
In addition, the potassium trifluoroborate salt **18** was synthesized. To generate the trifluoroborate salt, the pinacolboronate ester **5** synthesized previously was treated with potassium bifluoride in methanol resulting in an 88% yield of the salt **18** which was precipitated from the reaction mixture and was easily isolated by filtration and washing with organic solvents (Scheme 6).[135] This compound is stable to long-term storage and so was scaled up to 50 g for easier syntheses of future SMoC compounds. Suzuki reactions with this salt gave cleaner reactions with less of the deboronated side product which is difficult to separate from the reaction products. This is because deboronation occurs via the boronate compound which is in equilibrium with the boronic acid in solution.[129] Using the trifluoroborate salt will reduce the concentration of the borate in the reaction mixture and therefore suppress the deboronation reaction.



Scheme 6. Conversion of boronic ester to trifluoroborate salt. Reagents: Aqueous potassium bifluoride, MeOH, rt, 3 h (88%). Reaction carried out by Roberta Worthington.

2.2.1.3. *New Guanidinylation reagent*

A problematic step in the original SMoC synthesis was the addition of Boc-protected guanidine groups using the expensive *N*,*N*-di-Boc-*N'*-trifluoromethanesulfonyl-guanidine reagent, which resulted in average yields. To improve this step, a different guanidinylation reagent was used, *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine (Scheme 7). The amines of compound **6** were deprotected as described using a 33% solution of HBr in acetic acid and dried extensively under vacuum.[124] The dried amine was then suspended in a solution of DIEA in DCM and the guanidinylation reagent was subsequently added. On completion, the tetra-guanidinylated product was isolated by flash column chromatography. Using this reagent slightly increased the yield of guanidinylation from ~65% to over 70% at around 10% of the cost of the previous reagent. Therefore, the cost of SMoC production was significantly decreased, allowing for production of larger quantities for future studies.



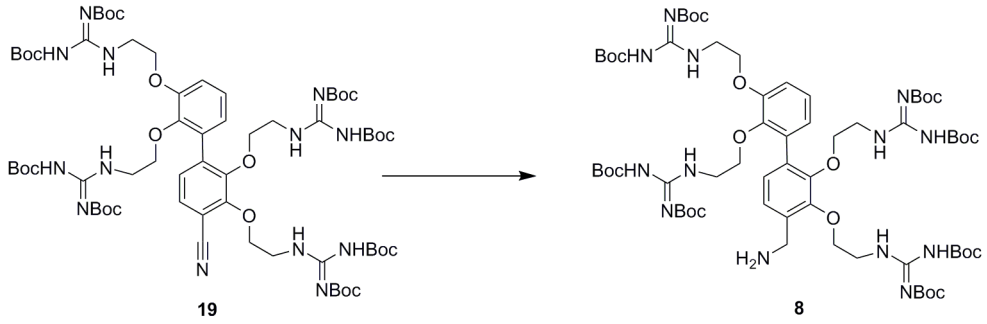
Scheme 7. Guanidinylation using pyrazole reagent. Reagents: a) 30% HBr in acetic acid, DCM, rt, 1h; b) *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine, DIEA, DCM, rt, 18 h (71%).

2.2.1.4. *Linker nitrile reduction*

NB. The work in this section was on going in the lab at the same time as the other experiments described in this chapter and was carried out by Roberta Worthington. It is included here for context as part of the overall development of the SMoC synthesis.

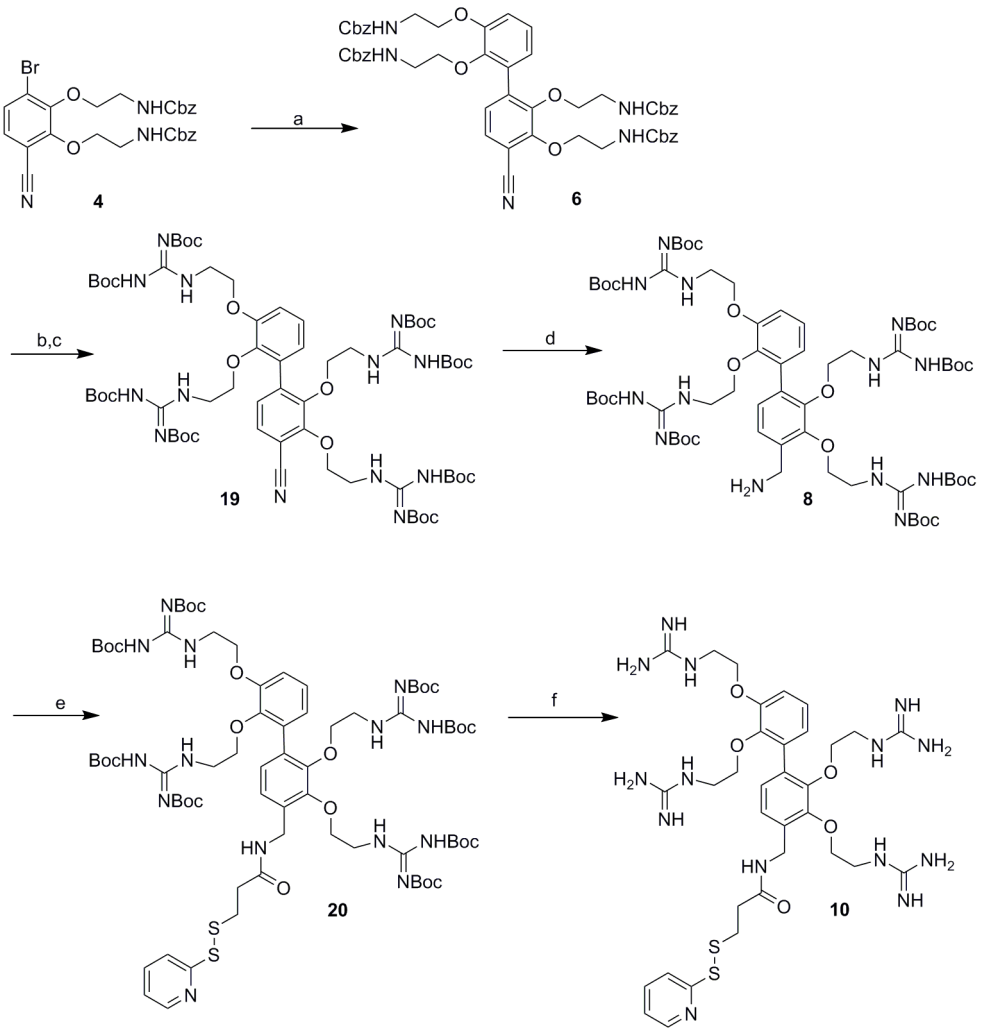
Finally, the reduction of the nitrile in order to produce a primary amine for attachment of a disulfide linker was another low-yielding step from the previous route. The previous synthesis used Raney Nickel/H_2 which resulted in 40-50%

yields. Reduction using sodium tetrahydroborate catalysed by cobalt chloride in a methanol solution, which has been shown to proceed via the formation of Co_2B which activates the nitrile to reduction by $NaBH_4$,[136] increased the yield of this step to ~60% (Scheme 8).[137]



Scheme 8. Reduction of nitrile to produce amine linker group. Reagents: $CoCl_2.6H_2O$, $NaBH_4$, MeOH, rt, 1 h (60%). Reaction carried out by Roberta Worthington.

The final optimised synthesis of 4G-SMoC-SSPy is therefore shown in Scheme 9. This improved method has enabled synthesis of 4G-SMoC-SSPy on a >200 mg scale for the first time.

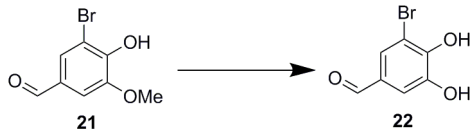


Scheme 9. New synthetic route to 4G-SMoC-SSPy. Reagents: a) **18**, $PdCl_2dppf.CH_2Cl_2$, Et_3N, $iPrOH/H_2O$, 82°C, 18 h (79%); b) 30% HBr in acetic acid, DCM, rt, 1h; c) *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine, DIEA, DCM, rt, 18 h (71%); d) $CoCl_2.6H_2O$, $NaBH_4$, MeOH, rt, 1 h (60%); e) 2,5-dioxopyrrolidine-1-yl-3-(pyridine-2-yldisulfanyl)-propanoate, DIEA, DCM, rt, 18 h (57%); f) TFA/m-cresol, DCM, rt, 3 h (82%).

2.2.2. 4G-SMoC with meta linker

In order to synthesize a 4G-SMoC derivative with a *meta* linker, the commercially available 5-bromovanillin **21** was used as the starting material, which contains the necessary functionality at the appropriate positions in order to form one half of the biphenyl structure. The 1-bromo group is required for the Suzuki-Miyaura coupling to the potassium trifluoroborate salt **18**, the hydroxy and methoxy groups are in the

correct positions for attachment of the guanidine side chains after demethylation of the methoxy group (Scheme 10), and the aldehyde *meta* to the bromo group supplies functionality in order to add a linker to this position. Demethylation of 5-bromovanillin was accomplished using $AlCl_3$ and pyridine followed by quenching with HCl,[138] and the catechol **22** was isolated in 80% yield after purification by flash column chromatography. The product was characterised by NMR, and complete demethylation was observed by the absence of a methyl peak (Figure 10).



Scheme 10. Demethylation of 5-bromovanillin. Reagents: $AlCl_3$, pyridine, chloroform, 0°C, 15 min; then warmed to 70°C, 17h (80%).

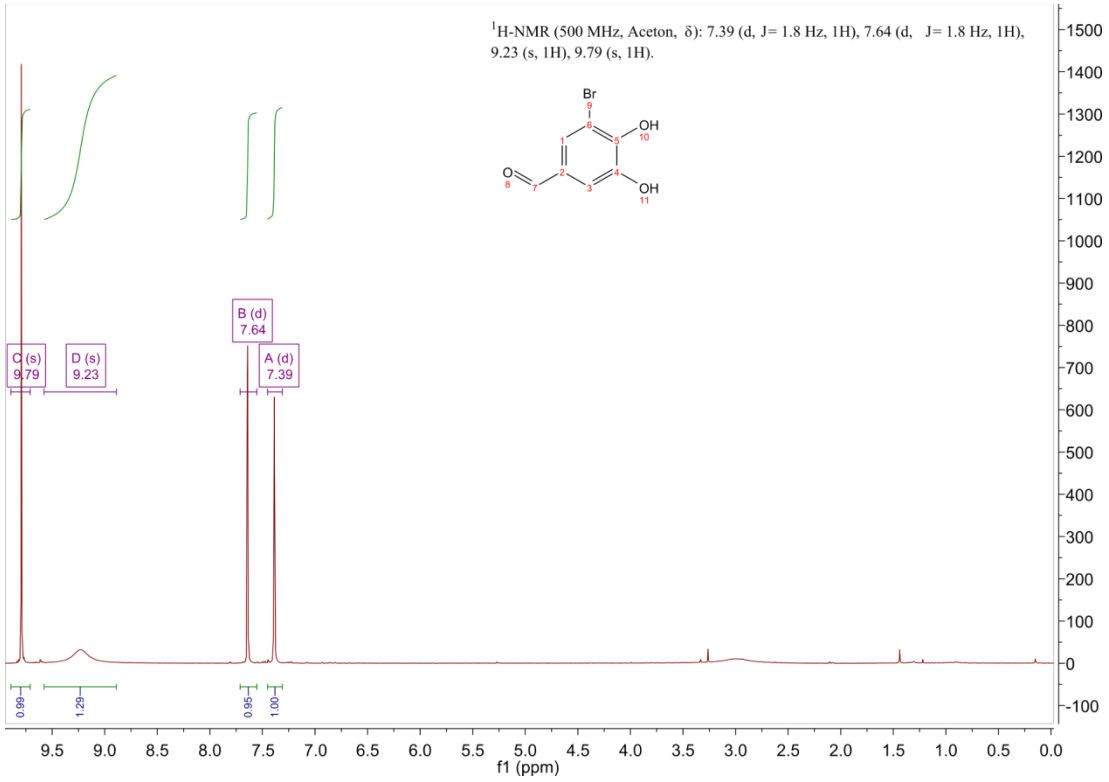
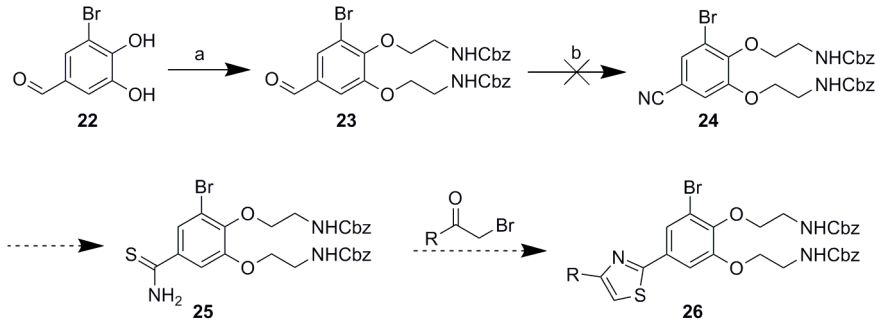


Figure 10. ^{1}H-NMR spectrum of catechol 22. There is no methyl peak in the alkyl region of the spectrum indicating complete demethylation. The peak at 7.39 ppm corresponds to the proton at position 3 and the peak at 7.64 corresponds to the proton at position 1 which is more deshielded due to its proximity to the bromo substituent. The broad peak at 9.23 ppm indicates the hydroxyl groups, and the peak at 9.79 ppm corresponds to the aldehyde proton.

2.2.2.1. *Thioamide linker*

In order to produce a functional linker at the *meta* position, the aldehyde group on the catechol was then transformed into a functional group onto which more complex structures could later to coupled. The first attempt to produce a *meta* linker focused on conversion of the aldehyde to a thioamide group (Scheme 11). This type of group has the ability to react with α-haloketones via the Hantzsch reaction, forming a thiazole ring linked to the required linker. This would enable attachment of various functional groups according to the requirements of the cargo. For example, attachment of the SSPy protein linker for reversible conjugation to proteins, or an acid group for non-reversible amide linkage to targeting peptides (Section 4.2.3).



Scheme 11. Proposed synthesis of thioamide linker. Reagents: a) **3**, Na_2CO_3, DMF, 80°C, 24 h (46%); b) See Table 6. R group represents the required linker functional group.

Alkylation of the catechol was carried out using the method described for 4G-SMoC using sodium carbonate in DMF at 80°C in order to add the Cbz-protected amino side chains to form **23** in 46% yield which was isolated by flash column chromatography. Since thioamides may be synthesized from nitriles via several different methods, the next attempted step was to convert the aldehyde to a nitrile using the nucleophile hydroxylamine to react with the aldehyde. In order to convert the aldehyde to the nitrile **24**, a one-pot synthetic method was used involving microwave irradiation of the aldehyde with hydroxylamine in NMP.[139] The first attempt at the reaction, heating in a microwave at 60°C resulted in 4 products, including the nitrile, oxime, and two unidentifiable products as detected by LC-MS. When the major product of the reaction was purified and analysed by high resolution mass spectrometry, it was found to be the oxime, suggesting that the OH group was not sufficiently activated (Figure). The reaction was repeated using temperatures of 80°C and 100°C, the same product was still formed, suggesting that temperature was not the limiting factor in the reaction. The reaction was repeated using DMF, another polar aprotic solvent with similar polarity to NMP, which gave the same product. The same reaction was attempted using the catechol product **22** rather than the alkylated compound, to establish whether the presence of the side chains affected the reaction. When heated in the microwave, the catechol product did not react, however when heated thermally, the nitrile product was formed and detected by LC-MS, but proved difficult to purify. Therefore, after several attempts at the aldehyde

to nitrile conversion (Table 6), the nitrile product could not be isolated at an acceptable purity and with sufficient yield. Although other strategies for the nitrile synthesis are possible, such as via the Schmidt reaction with hydrogen azide, due to time limitations, this line of enquiry was abandoned and an alternative linker strategy was sought.

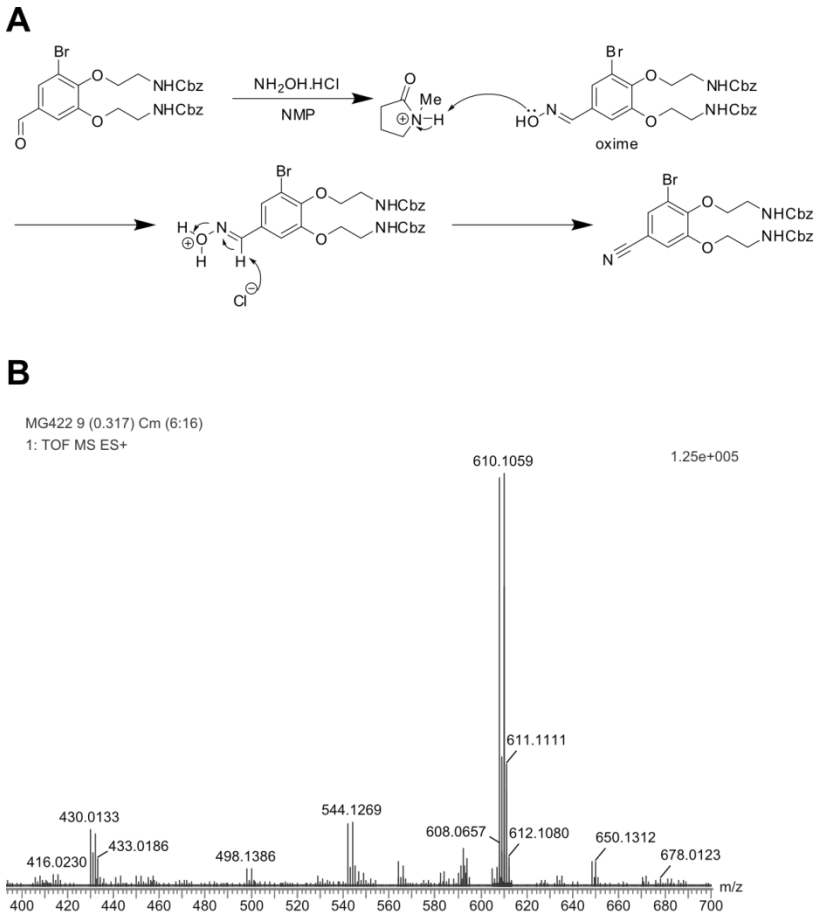
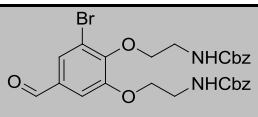
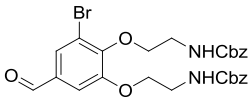
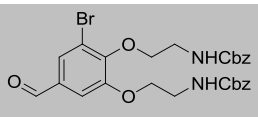
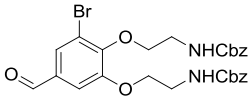
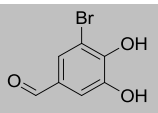
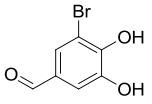


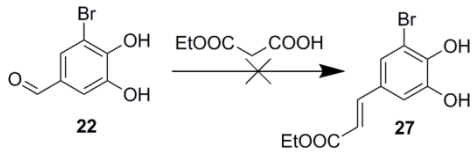
Figure 11. Conversion of aldehyde 23 to nitrile 24. A: Mechanism of the reaction. The oxime OH group must be activated by the transfer of a proton from NMP in order to form the water leaving group; B: High resolution mass spectrum of the major reaction product. The major peak at *m/z* 610.1059 corresponds $[M + Na]^+$ peak of the oxime (calculated *m/z* 610.1090), indicating inefficient activation of the oxime.

Table 6. Attempted nitrile 24 synthesis conditions.

Starting Material	Conditions	Result
	$NH_2OH.HCl$ (1.5eq) NMP 60°C (MW) 15 min	4 products (tlc). Oxime is major product (Figure)
	$NH_2OH.HCl$ (1.5eq) NMP 80°C (MW) 10 min	3 products (tlc). Oxime major product
	$NH_2OH.HCl$ (1.5eq) NMP 100 °C (MW) 5 min	3 products (tlc). Oxime major product
	$NH_2OH.HCl$ (1.5eq) DMF 100 °C (MW) 5 min	3 products (tlc). Oxime major product
	$NH_2OH.HCl$ (1.5eq) NMP 80 °C (MW) 5 min 90 °C (MW) 5 min 100 °C (MW) 15 min	No reaction observed (tlc)
	$NH_2OH.HCl$ (1.5eq) NMP 110 °C (thermal) 24 h	Product impure after flash column chromatography. Could not recrystallise.

2.2.2.2. *Acid Linker*

The next type of linker attempted was a carboxylic acid linker, which also provides sufficient versatility for attachment to macromolecules, such as proteins and peptides via amide coupling. The starting material used was the catechol product **22** synthesized previously, using the aldehyde to attach an ester which could be later hydrolysed to afford the acid linker group. A short hydrocarbon chain between the ring and the acid is preferable in order that steric effects between the guanidine groups and the SMoC cargo are limited. It was necessary to use an ester protecting group for the carboxylic acid to prevent the reaction of the carboxylate with the mesylate in the alkylation reaction, and also with the pyrazole guanidinylation reagent. The first attempt to add an ester group to form compound **27** used ethyl malonate with pyridine and piperidine which forms a carbon-carbon bond via an aldol reaction with the aldehyde (Scheme 12).



Scheme 12. Proposed aldol reaction to add ester group at linker position. Reagents: Ethyl malonate, pyridine/piperidine (1:0.03), rt, 17h.

The reaction yielded a mixture of products which could not be separated by flash column chromatography, and therefore an alternative strategy for the linker addition was sought (Scheme 13). Using a triphenylphosphine in DCM, the ester linker, separated from the ring system by an unsaturated carbon chain, was added via a Wittig reaction,[140] which completed successfully and the product **27** was purified and obtained in good yield (76%). Analysis of the *J* couplings of the double bond protons in the ^{1}H-NMR spectrum showed that the *E* stereoisomer was produced (*J* = 15.9 Hz, Figure 12).

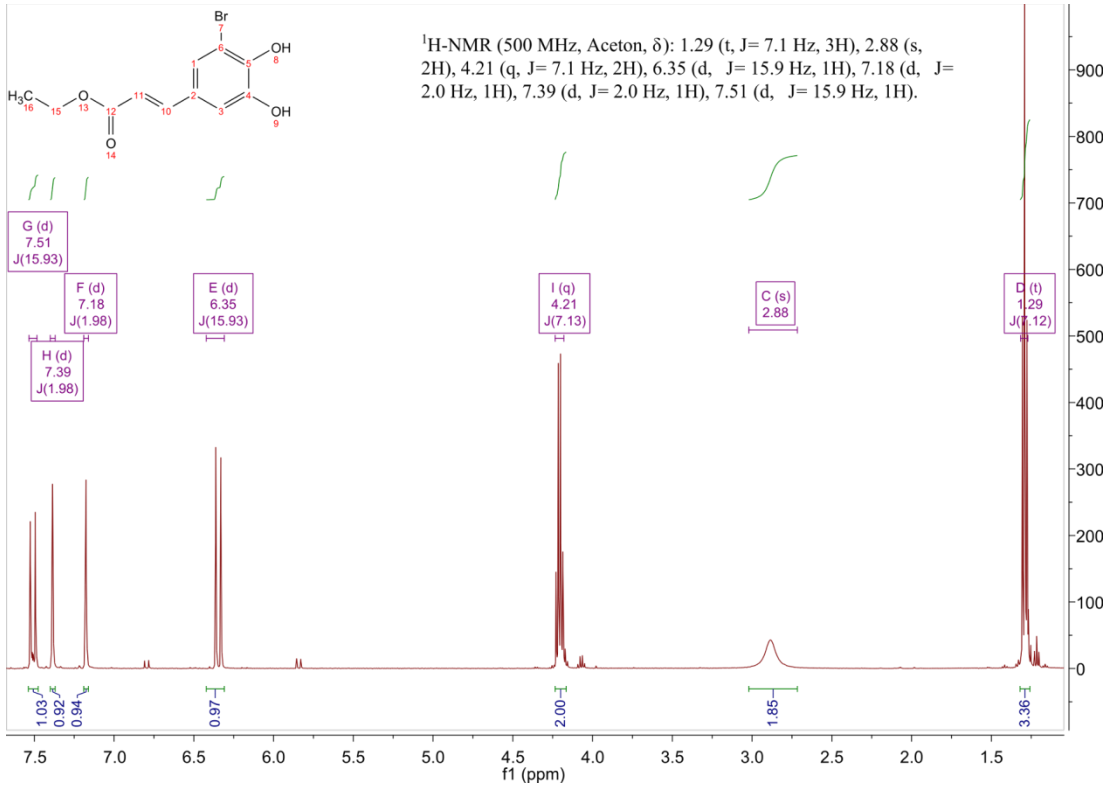
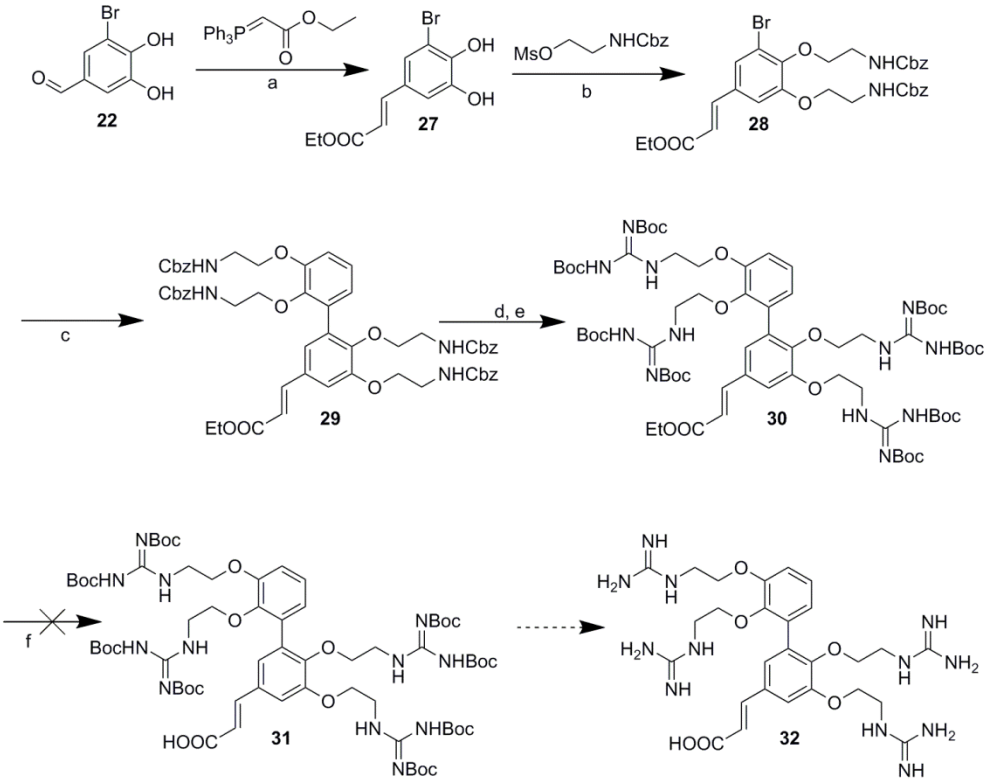


Figure 12. ^{1}H NMR spectrum of 27. The *J* coupling constant for the two double bond protons with chemical shifts (δ) 6.35 and 7.51 ppm is 15.9 Hz, indicating that the *E* stereoisomer is formed.

The catechol was alkylated via nucleophilic substitution with the mesylate compound as described above using caesium carbonate as the base at 65°C, which produced the alkylated product in near quantitative yield (96%) and isolated by flash column chromatography. The trifluoroborate salt synthesized previously was coupled to the ester via a Pd(0)-catalysed Suzuki-Miyaura coupling with triethylamine in an isopropanol/water solvent mixture (Scheme 13), which was heated in a microwave at 80°C, shortening the reaction time from 18 h to 10 min with very good yield (85%), and the coupled product was isolated by flash column chromatography. Following deprotection of the amines using 33% HBr in acetic acid, and thorough drying of the crude product, the side chains were guanidinylated with the new pyrazole reagent in DCM and DIEA, which resulted in a 55% yield of the Boc-protected product following isolation of the tetra-guanidinylated compound by flash column chromatography. The next step attempted was hydrolysis of the ester **30** to form the

acid linker (Table 7), which had been protected as the ester in order that the acid did not react with the guanidinylation reagent in the previous step. A solution of NaOH in methanol/dichloromethane was used to hydrolyse the ester, however no reaction was observed. The solvent was changed to methanol, and the methyl ester was formed due to the acid immediately reacting with the solvent. An aprotic solvent, 1,4-dioxane was tried which resulted very low conversion as observed by LC-MS. Finally, a stronger base, lithium hydroxide, was used and was also unsuccessful.



Scheme 13. Proposed synthetic strategy for the preparation 4G-SMoC with a meta carboxylic acid linker. Reagents: a) (Carbethoxymethylene)triphenylphosphorane, DCM, rt, 17h (76%); b) **3**, Cs_2CO_3, acetone, 65°C, 17h (96%); c) **18**, $PdCl_2dppf.CH_2Cl_2$, Et_3N, iPrOH/H_2O, MW 80°C, 10 min (85%); d) 30% HBr in acetic acid, DCM, rt, 1h; e) *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine, DIEA, DCM, rt, 18 h (55%); f) See Table 7 for conditions.

It was hypothesized that the stability of the ester was due to conjugation of the carbonyl with the adjacent double bond which was formed in the Wittig reaction. Due to time constraints, this line of enquiry was halted, however the final product may be obtainable via hydrogenation of the double bond, which may allow the ester to be hydrolysed under standard conditions.

Table 7. Attempts to hydrolyse ester 30.

Base	Conditions	Result
NaOH (2 eq)	MeOH/DCM (1:9) rt, 4h	No reaction (LC/MS)
NaOH (6 eq)	MeOH/DCM (1:9) rt, 4h	No reaction (LC/MS)
NaOH (0.5 M)	MeOH 50°C 5h	Methyl ester formed due to reaction of acid with methanol solvent.
NaOH (0.5 M)	Dioxane 50°C, 5h	Very low amount of product formed (LC/MS)
LiOH (5 eq)	THF/H_2O Base added at 0°, then warmed to rt.	No reaction (LC/MS)

2.2.3. 4G-SMoC disulfide-linked dimer

MALDI analysis of 4G-SMoC-SSPy **10** found that due to long-term storage, some of the compound had dimerised via disulfide linkages (Figure 13). It was therefore unclear whether the active species was the original 4G-SMoC, or this new disulfide compound.

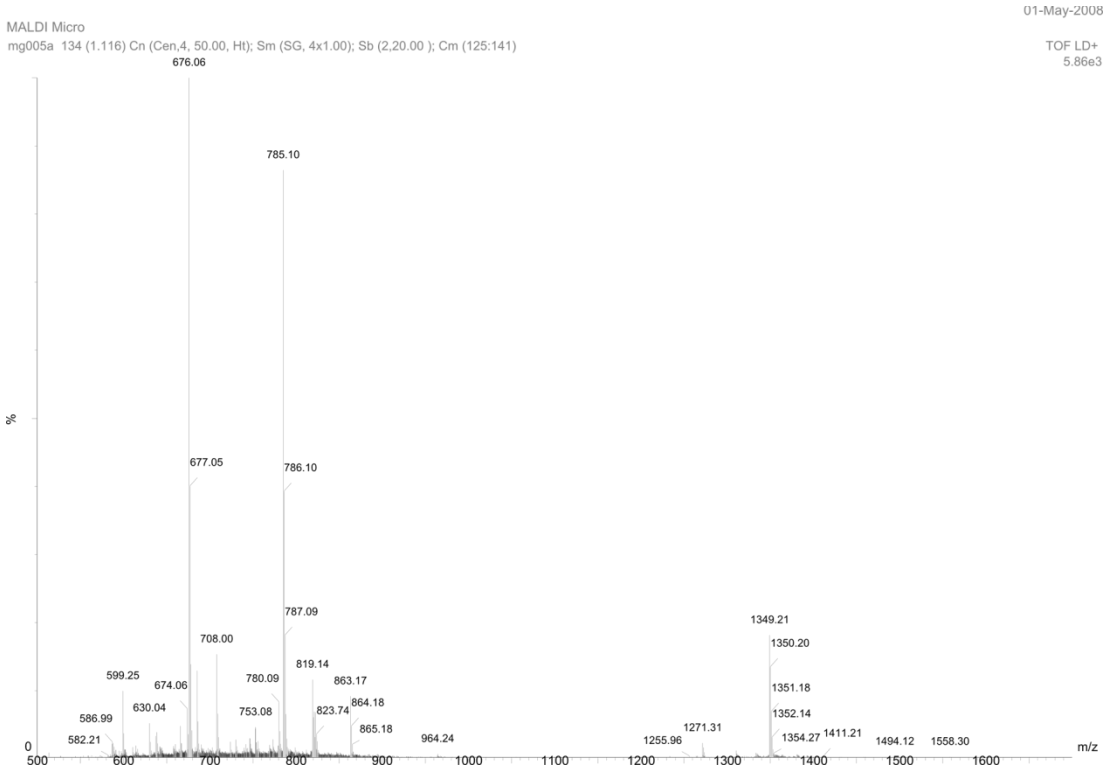
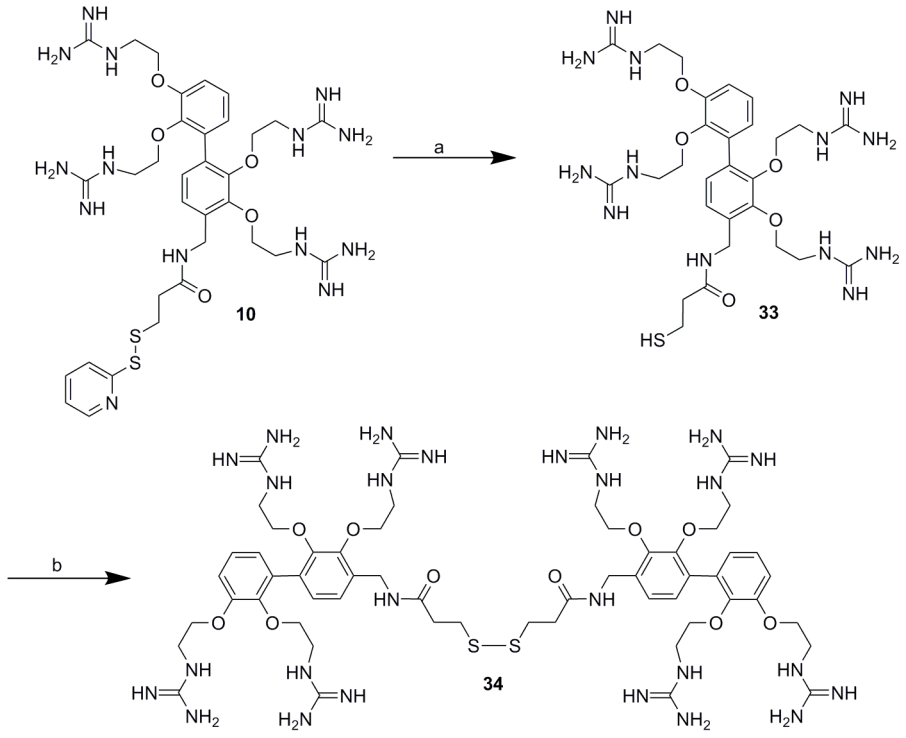


Figure 13. MALDI mass spectrum of 4G-SMoC-SSPy after long-term storage. The spectrum shows peaks at *m/z* 785.10 corresponding to 4G-SMoC-SPy, 676.06 corresponding to the thiol, and 1349.21 corresponding to the disulfide-linked dimer.

It was therefore decided to synthesize and purify the disulfide compound in order to establish the most effective compound for cell transfection. This was accomplished by reduction of the pyridyl compound with dithiothreitol (DTT) in methanol/water, followed by purification of the resulting thiol **33** by preparatory LC-MS. The thiol was then dimerised by stirring in an ammonium acetate solution for three days to form the disulfide **34** (Scheme 14) which was obtained in 55% yield after washing with ether. These reactions were monitored by LC/MS as shown in Figure 14. As only the +2 ion could be seen by electrospray ionisation, the final disulfide compound was further characterised by MALDI-TOF mass spectrometry in order to confirm that the conjugated product had been formed (Figure 15).



Scheme 14. Synthesis of 4G-SMoC-disulfide. Reagents: a) DTT, H_2O/MeOH (19:1), rt, 30 min; b) Ammonium acetate, H_2O, rt, 72 h (55% over 2 steps).

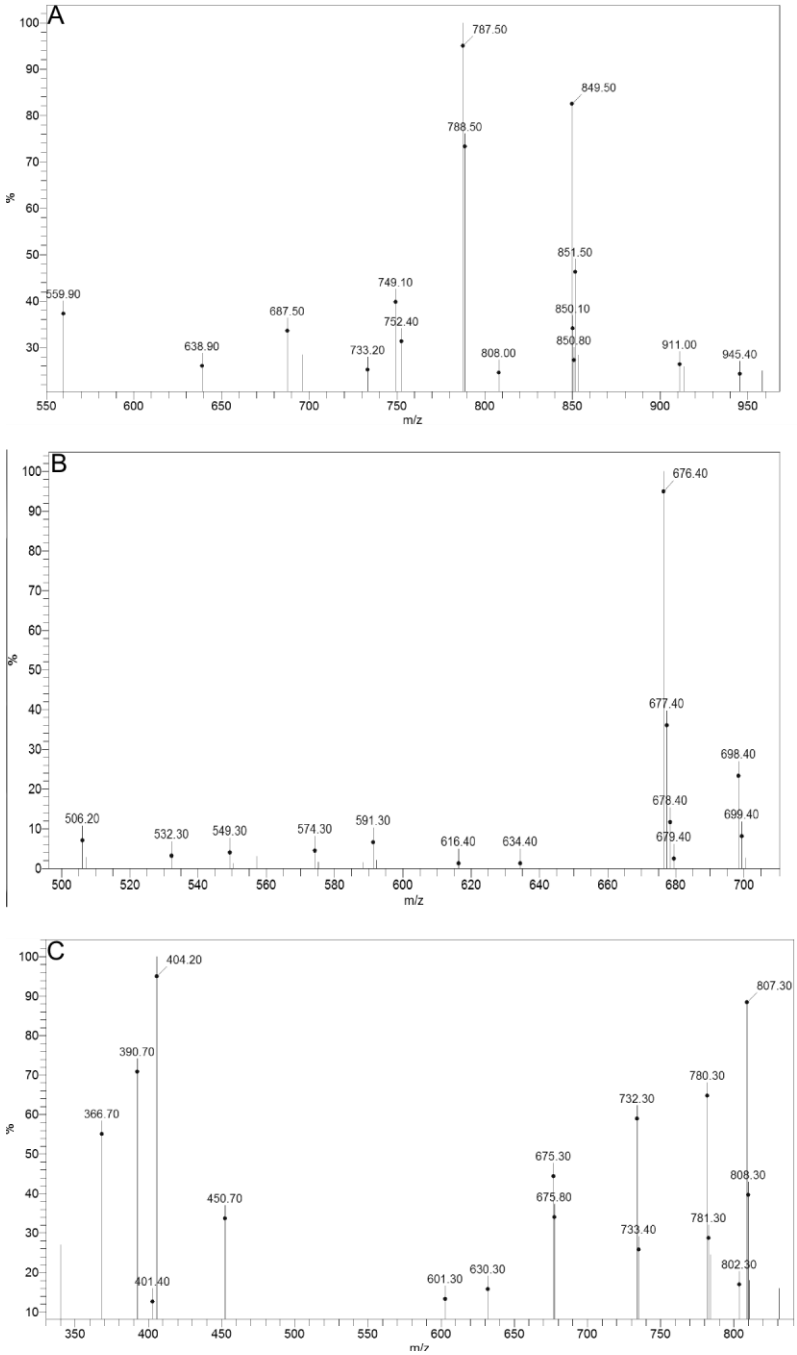


Figure 14. Monitoring of 4G-SMoC-disulfide formation by mass spectrometry. A: 4G-SMoC pyridyl starting material **10**. $[M + H]^+$ (*m/z* 786.50) $[M + H + 2CH_3OH)^+$ (*m/z* 849.50) ions observed ; B: After stirring with DTT for 30 min, the thiol **33** was formed. $[M + H]^+$ (*m/z* 676.40) and $[M + Na]^+$ (*m/z* 698.40) ions observed; C: After stirring in 7 g/l ammonium acetate solution for 72 h disulfide product **34** formed. $[M + 2H]^{2+}$ (*m/z* 675.80) ion observed.

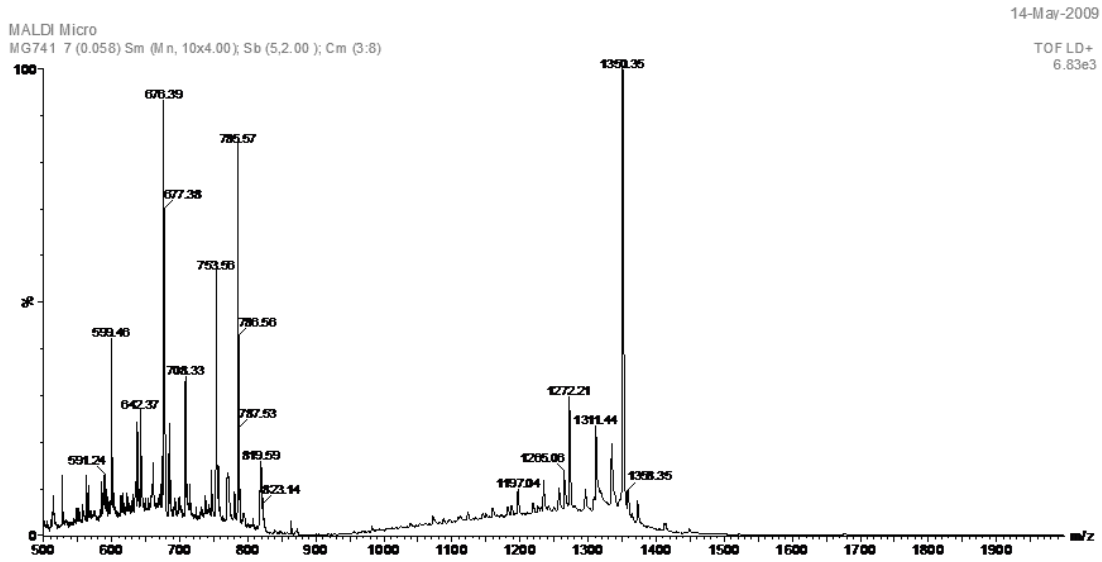
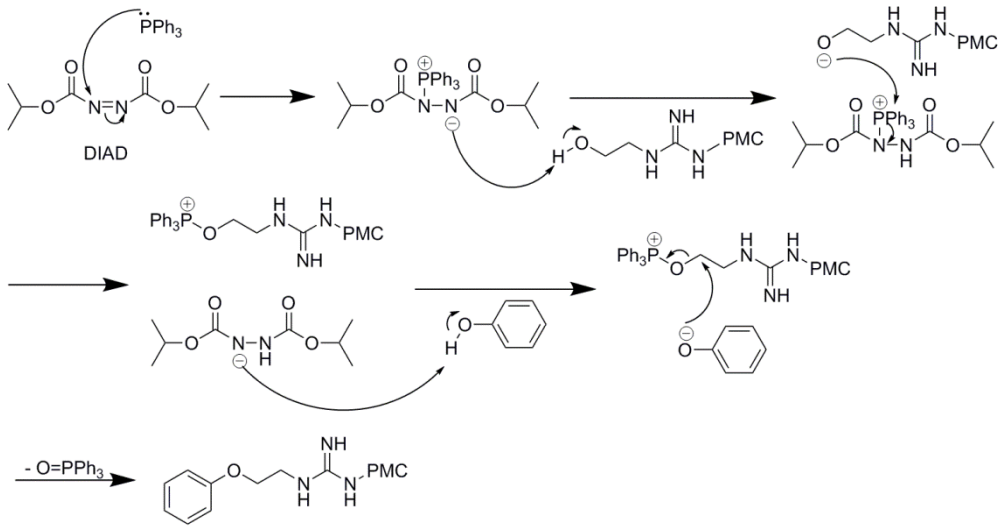


Figure 15. MALDI-TOF spectrum of 4G-SMoC dimer 34. $[M + H]^+$ (*m/z* 1350.35) ion observed.

2.3. Discussion

In this chapter, the synthesis of 4G-SMoC-SSPy has been examined and some steps were successfully optimised, allowing for a greater yield of product. However, it was not possible to reduce the number of steps by introducing PMC-protected guanidine groups to the catechol precursor using alkyation reagent **13**. The initial attempt to synthesize the mesylate derivative failed due to an intramolecular cyclisation resulting from the attack of the nucleophilic imine on the product. The Mitsunobu reaction was then attempted in order to couple **13** directly to phenols (Scheme 15). The mechanism of this reaction involves initial attack by a phosphine on the azo-containing Mitsunobu reagent to form a basic zwitterionic intermediate which is able to deprotonate the alcohol, forming an alkoxide. This then attacks the positively charged phosphorus forming an oxophosphonium species and a second basic nitrogen on the Mitsunobu reagent which can deprotonate the nucleophilic starting material, in this case a phenol. The deprotonated phenol then attacks the carbon of the oxophosphonium derivative, with phosphine oxide as the leaving group to form the coupled product.



Scheme 15. Mechanism for coupling of the guanidine-PMC sidechain to phenol via a Mitsunobu reaction using DIAD/TPP.

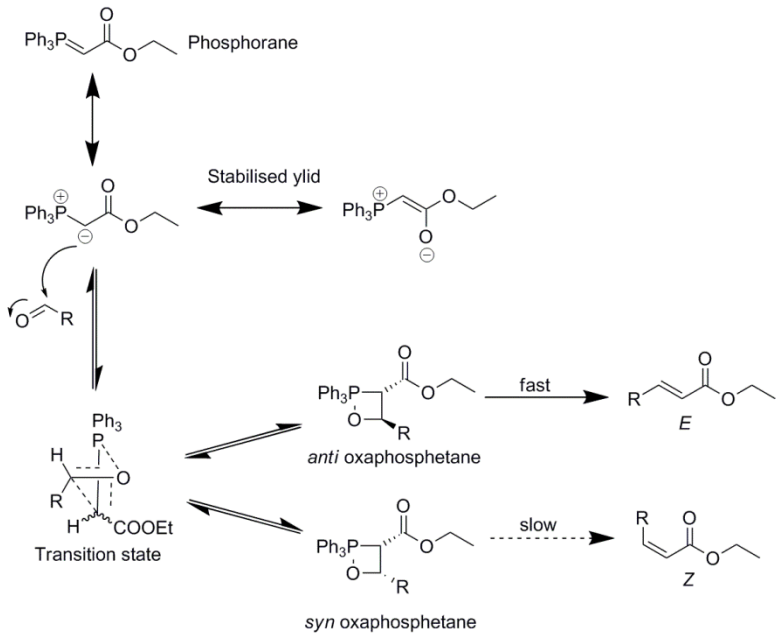
Despite the Mitsunobu reaction being a widely reported method for the alkylation of phenols, in this case the reaction failed to proceed, even when using harsh reagents. Using more acidic phenols by adding electron withdrawing groups to the ring may allow this reaction to proceed.

Despite this result, the guanidinylation and nitrile reduction steps have been improved by using different reagents which allow a higher yield of product at lower cost. In addition, a new trifluoroborate Suzuki coupling reagent has been synthesized in high yields which is highly stable to storage and reacts with fewer side reactions than the previous boronic ester. Thus the aim of optimising the 4G-SMoC synthetic route has been successful and has allowed a larger quantity of 4G-SMoC-SSPy to be synthesized than with the previously described method. In addition, the synthesis of future SMoCs has been made more efficient by the stable trifluoroborate intermediate.

The aim of finding a synthetic route to a 4G-SMoC derivative with a *meta* linker has also been partially successful. A key catechol intermediate with an aldehyde at the required linker position was successfully synthesized in good yield, which contains

the necessary functionality at the appropriate positions to form one half of the SMoC biphenyl structure. This intermediate may be successfully alkylated to add the amino side chains, and the bromine successfully participates in a Suzuki-Miyaura coupling reaction with the trifluoroborate intermediate to form the biphenyl backbone. The attempt to produce a thioamide linker was unsuccessful, due to the difficulty in forming a nitrile at the linker position, however an ester derivative was successfully synthesized using a Wittig reaction. The resulting product was the *E*-stereoisomer, as expected when using a phosphorane compound containing a carbonyl which is able to form a stabilised ylide intermediate. The stabilised ylide means that subsequent formation of the oxaphosphetene is reversible, and therefore the reaction is thermodynamically controlled, proceeding via the more stable *anti* diastereomer to form the *E*-alkene.[141] This is in contrast to using an unstabilised ylide, where the reaction is kinetically controlled. The oxaphosphetane formation is irreversible, and therefore the *syn* diastereomer is produced via the most stable transition state, which leads to *Z*-alkene product (Scheme 16). The Boc-protected SMoC intermediate was successfully synthesized using the optimised reactions described previously.

Ester hydrolysis was unsuccessful under the attempted conditions, which may be due to conjugation with the adjacent double bond. It has been observed previously that the hydrolysis of unsaturated esters is considerably slower than expected from the predicted steric effects of the substituents.[142] When the unsaturation is moved further from the ester group, this effect decreases suggesting that unsaturation is responsible for the lower rate of reaction. With further experimentation, such as hydrogenation of the double bond, it is possible that this intermediate may be hydrolysed, and converted to the required SMoC compound by cleavage of the Boc protecting groups using the previously described method.



Scheme 16. Mechanism of the Wittig reaction. The starting phosphorane is a canonical form of the ylide, which is stabilised by conjugation with the carbonyl group. Nucleophilic attack by the ylide onto the aldehyde can proceed via two possible transition states. The more stable transition state, where the bulky groups of the ylide and the aldehyde are furthest apart, leads to the *syn* oxaphosphetane, whereas the less stable transition state forms the *anti* oxaphosphetane, which is more stable since the bulky groups are on opposite sides of the molecule. Since the ylide is stabilised, the formation of the oxaphosphetane is reversible, and so the reaction is under thermodynamic control, favouring formation of the *anti* diastereomer, and hence the *E*-alkene product.

Finally, a SMoC dimer consisting of two 4G-SMoC molecules joined via a disulfide linkage has been synthesized and purified. This compound may be used in cell uptake assays to compare its internalisation efficiency to 4G-SMoC-SSPy in order to determine the active SMoC compound. Due to the increased number of guanidines, the dimer compound may have a higher affinity for negatively charged proteoglycans and phospholipids and therefore may show a greater internalisation efficiency than the standard 4G-SMoC compound.

CHAPTER 3

Exploring the interactions of SMoCs with siRNA

3. Exploring the interactions of SMoCs with siRNA

3.1. Introduction

Given that SMoCs have been shown to be successful in delivering active proteins into cells, another useful application would be the delivery of siRNA, a class of macromolecule with great therapeutic potential. CPPs that have been used for siRNA delivery are either covalently attached to siRNA or packaged as electrostatic complexes, making use of their cationic arginine and lysine residues. Since SMoCs are highly positively charged due to the presence of guanidine groups, it is hypothesized that they may form non-covalent complexes with siRNA which are capable of crossing the cell membrane resulting in gene-specific knockdown.

3.1.1. The SMoC π-cation interaction

Molecular dynamics studies on penetratin have suggested that the aromatic tryptophan and phenylalanine amino acids form π-cation interactions with arginine residues in the same molecule during attachment to the cell membrane.[143] It was suggested that these interactions help to shield the positive charges from the water phase and therefore enhance insertion into lipid bilayers.

π-Cation interactions were first described by Burley and Petsko in 1986 when they observed that the side chains of cationic amino acids such as lysine and arginine are preferentially located near the π-electron cloud of aromatic amino acids such as phenylalanine, tyrosine and tryptophan.[144] It has been found that π-cation interactions are very common in proteins, with one interaction per 77 residues on average. Arginine residues are more favoured for π-cation interactions than lysine, probably due to the guanidine group on arginine being larger and less solvated, resulting in greater van der Waals interactions with aromatic rings than that of lysine. In addition, guanidines have the ability to form hydrogen bonds and π-cation interactions simultaneously, whereas the lysine sidechain can only participate in one type of binding at a time.[145] Guanidines can bind in two possible geometries, "T-shaped" and "parallel", with T-shaped being favoured in the gas phase and parallel

favoured in solution, although T-shaped interactions are stronger on average (-10.6 kcal/mol for T-shaped compared to -4.1 kcal/mol for parallel).[145]

Since SMoCs are amphipathic, containing a hydrophobic biphenyl backbone furnished with charged guanidine side chains, it is hypothesised that the guanidine groups may interact with the biphenyl structure via π-cation interactions. Such interactions may therefore enhance membrane insertion in the same way as in penetratin. In addition, it is hypothesised that π-cation interactions in SMoCs may also stabilise the guanidine positive charges, raising the guanidine pK_a, therefore allowing SMoCs to interact more readily with anions than free guanidine groups. This may have implications for the binding of SMoCs to siRNA as well as to cell surface proteoglycans, which are thought to bind CPP arginine side chains, leading to internalisation by endocytosis (Section 1.2.1.1).

In order to test this hypothesis, the presence of π-cation interactions in SMoCs may be predicted using molecular dynamics computational simulations and the magnitude of any interaction may be measured by comparing the pK_a values of different SMoC structures determined using NMR spectroscopy.

3.1.2. Characterisation of SMoC-siRNA complexes

It is hypothesized that siRNA may be delivered to the cytoplasm using SMoC-siRNA electrostatic complexes, as observed for several CPPs. To test this, such complexes should be characterised, including determining the binding affinities of different SMoCs for siRNA. Electrostatic interactions with oligonucleotides are often characterised using a gel shift assay, in which the oligonucleotide is mixed with increasing concentrations of the counter-ion species and the resultant complexes run on an agarose gel.[108, 111, 117] Agarose gel electrophoresis involves the migration of negatively charged oligonucleotides through a gel towards a positive electrode. The oligonucleotides on the gel are detected using the dye ethidium bromide (EtBr) which fluoresces when bound between the base pairs of oligonucleotide molecules. When the oligonucleotide is bound to a cation, the overall charge is decreased, and

hence the speed of migration through the gel is retarded. Therefore, addition of increasing levels of SMoC to siRNA is predicted to cause the siRNA band to shift up the gel, until the point at which the siRNA charge is saturated when the band will remain static. Since the siRNA and SMoC components are at equilibrium, it may therefore be possible to calculate a binding affinity based on the amount of free siRNA that is unbound to SMoCs. The binding of various SMoCs may be compared against oligoarginine peptides which have been shown to complex siRNA for delivery into cells. If the π-cation stabilisation of the SMoC guanidine groups is sufficient to affect binding to siRNA, it is predicted that 4G-SMoC will bind siRNA with a greater affinity than tetraarginine (R4).

In order to gain a more accurate measure of binding affinity of siRNA to SMoCs, a more quantitative technique should be used, such as isothermal titration calorimetry (ITC). The isothermal titration calorimeter measures the temperature change when small aliquots of a ligand are added to a macromolecule contained in a sample cell. It does this by measuring the power difference required to maintain a constant temperature between the sample cell and a reference cell containing only buffer. By integrating the resulting spikes corresponding to the temperature evolved upon each addition of ligand, a binding curve may be produced, and hence the binding affinity of the ligand to the macromolecule may be calculated. This method can produce very accurate values for binding affinity, stoichiometry and enthalpy changes. A new SMoC compound, 6G-SMoC, containing six guanidine groups attached to a terphenyl backbone with no linker will be synthesized and is hypothesised to have a greater binding affinity than the existing 4G-SMoC-SSPy. The omission of a linker group is predicted to reduce steric hindrance to siRNA binding, as no linker is required in the formation of SMoC-siRNA complexes.

3.1.3. SMoC-mediated siRNA transfection

Following formation of electrostatic SMoC-siRNA complexes, it is necassary to test whether these structures are capable of crossing the cell membrane. siRNA linked to a fluorescent dye such as Cy3 may be used to determine uptake efficiency by FACS

analysis, or for visualisation of transfected cells using confocal microscopy. These methods are useful for determining whether a specific delivery vector is effective at promoting cellular uptake, however they do not assess whether the siRNA is active once inside the cell and able to knockdown its target mRNA. Experiments resulting in entrapment in endosomes, or degradation of the siRNA may give positive results in these uptake assays, but unless the siRNA is able to successfully mediate RNAi, the transfection cannot be considered successful. Therefore, it is important that siRNA uptake results are accompanied by functional data measuring the knockdown of the target gene. This may be achieved by such methods as quantification of the target mRNA in transfected and untreated cells, Western blot analysis of the target protein, or measurement of the activity of the target protein. In order to assess the potential for the use of SMoCs as siRNA transfection agents, uptake data as well as functional data should be acquired in order to determine the efficacy of both internalisation and knockdown.

In this study, quantitative real-time PCR (qRT-PCR) will be used to measure the extent of mRNA knockdown caused by siRNA transfection. A one-step qRT-PCR kit will be used which includes a reverse transcriptase, which synthesizes the corresponding cDNA strands from the mRNA contained in each sample prior to amplification, followed by ~40 cycles of heating and cooling to amplify the required DNA sequence. During each PCR cycle, the temperature of the reactions is raised to 95°C to denature the DNA strands, followed by cooling to 60°C which enables annealing of the PCR primers to the target sequences of the separated strands and the *Taq* polymerase enzyme extends the strand using the complementary DNA as a template. In qRT-PCR, a dye such as SYBR Green is used which emits fluorescence when bound to double stranded DNA. The amount of dsDNA corresponding to the required gene will double on each cycle, and therefore the fluorescence of the solution will increase exponentially (Figure 16A). By comparing the cycle number at which the fluorescence of each sample raises above background noise, crossing the so-called cycle threshold (C_t), the relative amounts of template can therefore be calculated **(**Figure 16B). A housekeeping gene, such as GAPDH, is used to

normalise the mRNA content in each sample. The percentage knockdown of mRNA can be calculated by comparing cells transfected with the target siRNA to those treated with a non-targeting siRNA in order to take into account any off-target effects, or side effects caused by the transfection reagent. It is hypothesized that treatment of cultured cells with SMoC-siRNA complexes will result in knockdown of the target gene as a result of cell internalisation. Lipofectamine, a standard cationic liposome-based transfection reagent, will be used as a control to determine whether successful knockdown has been acheived.

The overall aims of this chapter are therefore to determine the existance of an intramolecular π-cation interaction in SMoC molecules using molecular dynamics and NMR spectroscopy; to characterise the formation of SMoC-siRNA complexes using gel shift assays and ITC; and to test whether siRNA may be delivered to the intracellular environment of cultured cells by measuring mRNA levels using qRT-PCR.

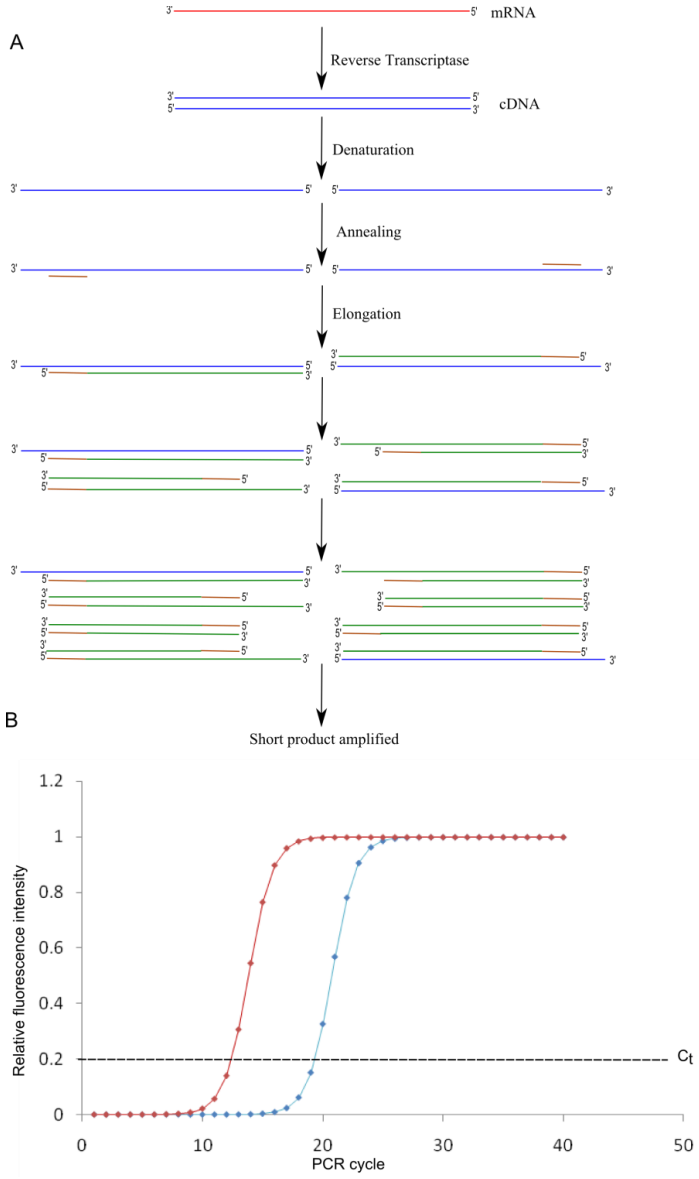


Figure 16. qRT-PCR mechanism. A: Initially, mRNA extracted from the cells is converted to cDNA using a reverse transcriptase enzyme. The cDNA is then denatured by heating to 95°C, allowing the strands to separate. The mixture is then cooled to 60°C to allow the PCR primers for the chosen gene to anneal to their complementary sequences on the template. *Taq* polymerase then extends the primers by the addition of bases complementary to the template strand. As this is repeated, the DNA of the required gene is amplified exponentially. B: SYBR Green dye binds to dsDNA, and therefore can be used to quantitate the relative number of copies of the target gene in the original sample. This is calculated using the C_t value - the number of cycles required for the fluorescence to rise above the background noise level. Since the number of copies of the target gene doubles at each cycle during exponential growth, the difference in the C_t value between two samples, n, represents a 2^n difference in the number of original copies of the target gene.

3.2. Results

3.2.1. Molecular Dynamics

In order to gain more information about the structural conformation of the SMoC compounds, some simple SMoC-like structures were analysed using the molecular dynamics software YASARA under the AMBER99 force field (Figure 17). The simulations were carried out in vacuo in order to eliminate interference from the solvent. Although this may not give an accurate representation of how the compounds behave in solution, this experiment will confirm whether it is sterically favourable for π-cation interactions to form between the guanidines and the biphenyl structure, which was more fully investigated in solution by NMR spectroscopy. It was found that in the case of a 2G-SMoC, possessing two guanidine side chains at the *ortho* position to the biphenyl bond, the guanidine groups form T-shaped π-cation interactions with the opposite ring. This type of π-cation interaction is also seen in a 1G-SMoC structure with one guanidine side chain *ortho* to the biphenyl bond. However, when the side chain is moved to the meta position, the carbon chain is not long enough to allow interaction with the opposite ring, and a weaker parallel π-cation interaction is seen between the guanidine and the adjacent ring. When only a single phenyl ring is present, no interaction occurs, suggesting that the electron density of one ring is insufficient to promote a π-cation interaction. The lengths of the π-cation interactions found in the SMoC structures were measured from the centre of the ring to the guanidinium carbon atom (Table 8). The lengths of the T-shaped interactions are ~3.8-4.2 Å, whereas the parallel interactions are ~3.3 Å. These lengths are consistent with Gallivan and Dougherty's extensive study of π-cation interactions which suggests that 99% of all significant π-cation interactions in PDB proteins are ≤ 6.0 Å.[145]

These results suggest that the SMoC guanidinium ions may be stabilised by the presence of the phenyl ring, thus allowing stronger electrostatic interactions with anions.

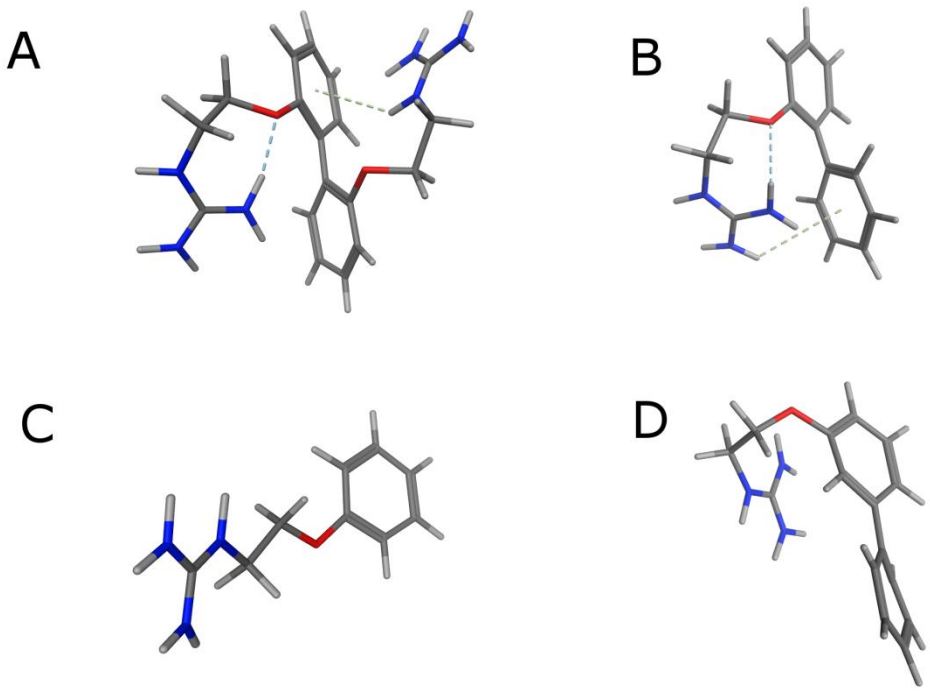
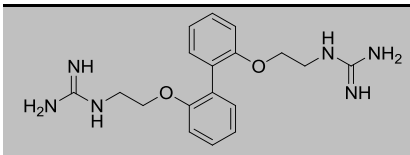
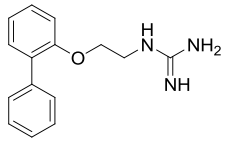
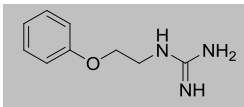
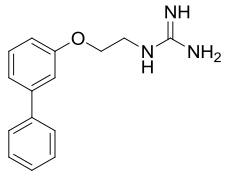


Figure 17. Snapshots of a molecular dynamics simulation of simple SMoC-like molecules containing one or two guanidine side chains. In a simple 2G-SMoC molecule (A), T-shaped π-cation interactions form between the side chains and their opposite rings. This is also the case when one side chain is removed (B). For a mono-phenyl structure with one side chain (C), no interaction is present, whereas the addition of a second ring *meta* to the side chain (D) promotes a parallel π-cation interaction onto the adjacent ring.

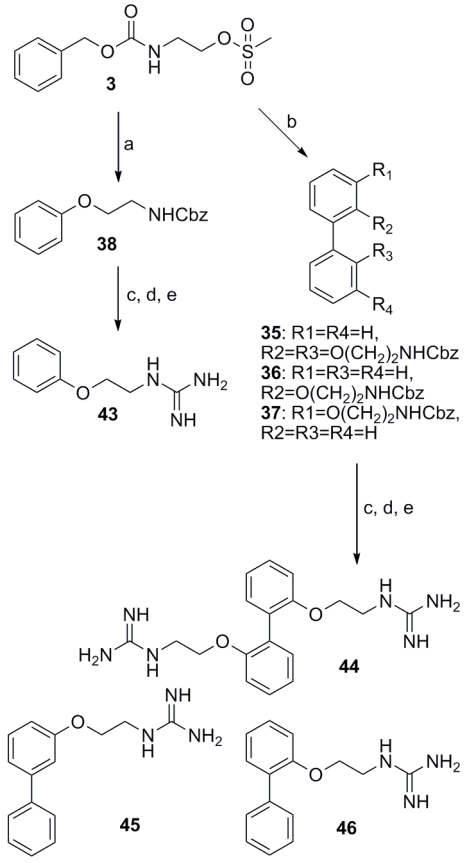
Table 8. Measured lengths of the π-cation interactions observed in 1G- and 2G-SMoC structures.

Structure	Type of interaction	Distance (Å)
	T-shaped	3.8-4.2
	T-shaped	3.8
	None	
	Parallel	3.3

3.2.2. NMR pK_a study

Having shown the theoretical possibility of a π-cation interaction in SMoCs using molecular dynamics, the detection of such interactions was then attempted using NMR spectroscopy. For this experiment, the simple 1G- or 2G-SMoC derivatives that were analysed by molecular dynamics were synthesized (Scheme 17) from commercially available mono- or biphenyl starting materials, eliminating the Suzuki coupling step. The hydroxyl groups were alkylated using the mesylate reagent **3**, which was synthesized as described,[124] using $CsCO_3$ in acetone at 65ºC and the products purified by flash column chromatography to afford the dialkylated product in 35% yield, and the monoalkylated products in 70-100% yields. The amines were then deprotected using 33% HBr in acetic acid, followed by extensive drying of the crude products under vacuum. The new pyrazole guanidinylation reagent with DIEA was used to add Boc-protected guanidine groups, which were obtained by

purification by flash column chromatography in 26-51% yields. The Boc groups were subsequently deprotected in a solution of TFA using TIPS as a cation scavenger, to give the four new SMoC compounds in quantitative yields after washing with ether. The final compounds were characterised by NMR, in order to determine which peaks to observe for the measurement of pK_a (Figure 18).



Scheme 17. Synthesis of simple SMoC compounds for determination of guanidine pK_a. Reagents: a) Phenol, Cs_2CO_3, acetone, 17h, 65°C; b) 2-, 3-, or 2,2'- dihydroxybiphenyl, Cs_2CO_3, acetone, 65°C, 17h; c) 30% HBr in acetic acid, DCM, rt, 3h; d) *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine, DIEA, MeCN, rt, 48h; e) TFA/TIPS/H_2O, rt, 4h.

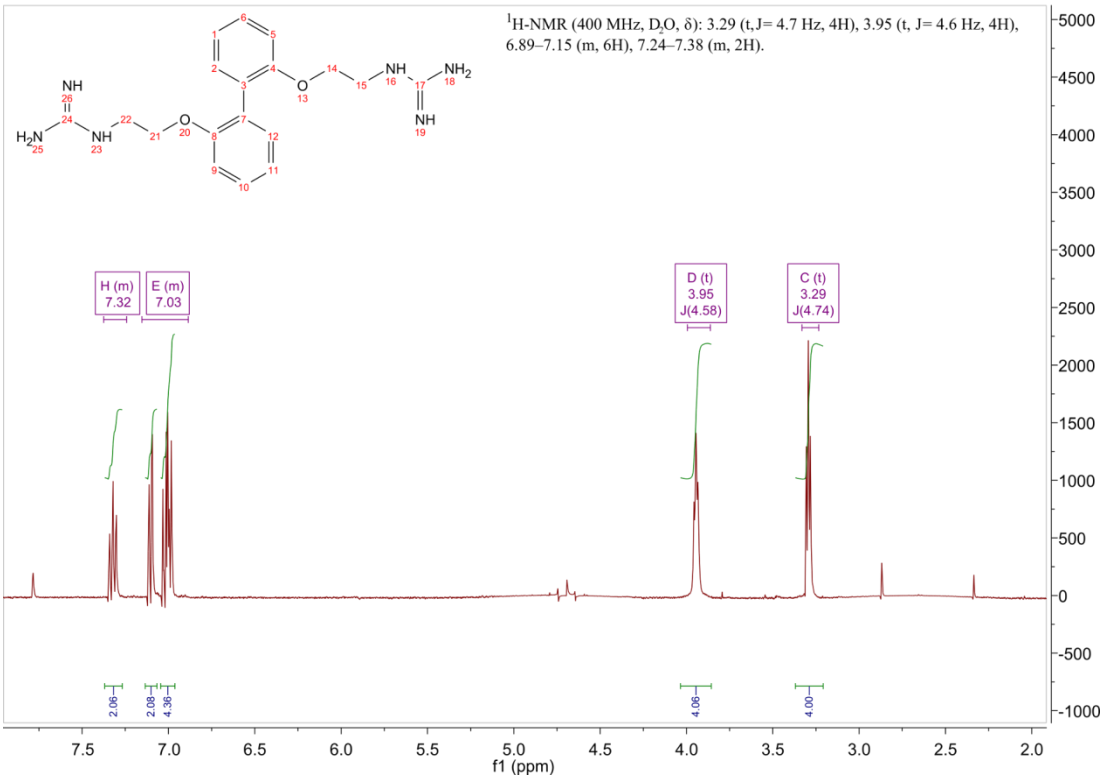


Figure 18. 1H-NMR spectrum of 2G-SMoC 44. The triplet at 3.29 ppm corresponds to the methylene protons adjacent to the guanidine group; the triplet at 3.95 corresponds to the methylene protons adjacent to the ether oxygen. The protons in the ethyl group show spin coupling with a coupling constant *J* of 4.6-4.7 Hz. The multiplet at 7.03 ppm corresponds to six aromatic protons, and the multiplet at 7.32 ppm corresponds to the two protons adjacent to the biphenyl bond. Since the spectrum was measured in D_2O, free exchange of the guanidine protons with the solvent eliminated these peaks. The triplet at 3.29 ppm is the most appropriate peak to observe in order to measure pK_a, as this will be the most affected by deprotonation of the guanidine group.

Comparing the pK_a of the guanidines in the simple mono-phenyl compound **43** and the biphenyl compound **44** will show whether π-cation interactions take place between SMoC guanidine groups and their opposite phenyl ring and the strength of these interactions. Molecular dynamics studies have shown that when the guanidine side chain is moved to the position *meta* to the biphenyl bond, the alkyl chain is of insufficient length to allow T-shaped π-cation interactions onto the opposite ring, but that the weaker parallel π-cation interactions may occur between the guanidine and the adjacent ring. Therefore, for compound **45**, a lower pK_a may be expected than

for **46**. The strength of this weaker interaction may also be measured by comparing **45** with **43**.

For each compound, the change in ionisation state of the guanidine groups was measured using NMR by titrating each compound with sodium hydroxide and recording the ^{1}H chemical shifts of the CH_2 protons adjacent to the guanidine group against the pH of the solution (Figure 19). The pK_a of the guanidine group(s) is given by the Henderson-Hasselbalch equation

$$\mathrm{p}K_a = \mathrm{pH} + \log\frac{[BH^+]}{[B]}$$

Deprotonation of the guanidine(s) will have a significant effect on the chemical shift of the protons on the adjacent methylene group(s) due to the change in electron density around the nitrogen atom, resulting in increased shielding of the CH_2 protons. The same effect could have been observed by measuring the ^{13}C shift of the guanidine carbon atom, however this was rejected in favour of the proton NMR approach for practical reasons, including the more time-consuming nature of gathering ^{13}C-NMR data, and the higher concentration of compound needed to ensure a high quality spectrum. The chemical shift of the charged form (δB^+) may be determined by measuring the chemical shift at a low pH. For these compounds, the chemical shift was found to be at its maximum below pH 10. Likewise, at very high pH (pH 14), the chemical shift of the deprotonated guanidine groups may be measured (δB). Thus, the chemical shift at a given pH (δA) is representative of the proportion of groups that are protonated at a given time. Therefore, the equation may be expressed in terms of change in chemical shift:

$$p\mathrm{K}_a = \mathrm{pH} + \log\frac{\delta A - \delta B}{\delta B^+ - \delta A}$$

Hence, by plotting the log of the ratio of the charged and uncharged forms expressed as chemical shifts against the pH of the solution, the pK_a may be calculated from the y-intercept of the linear trend line (Figure 20). The calculated pK_a values are given in Table 9.

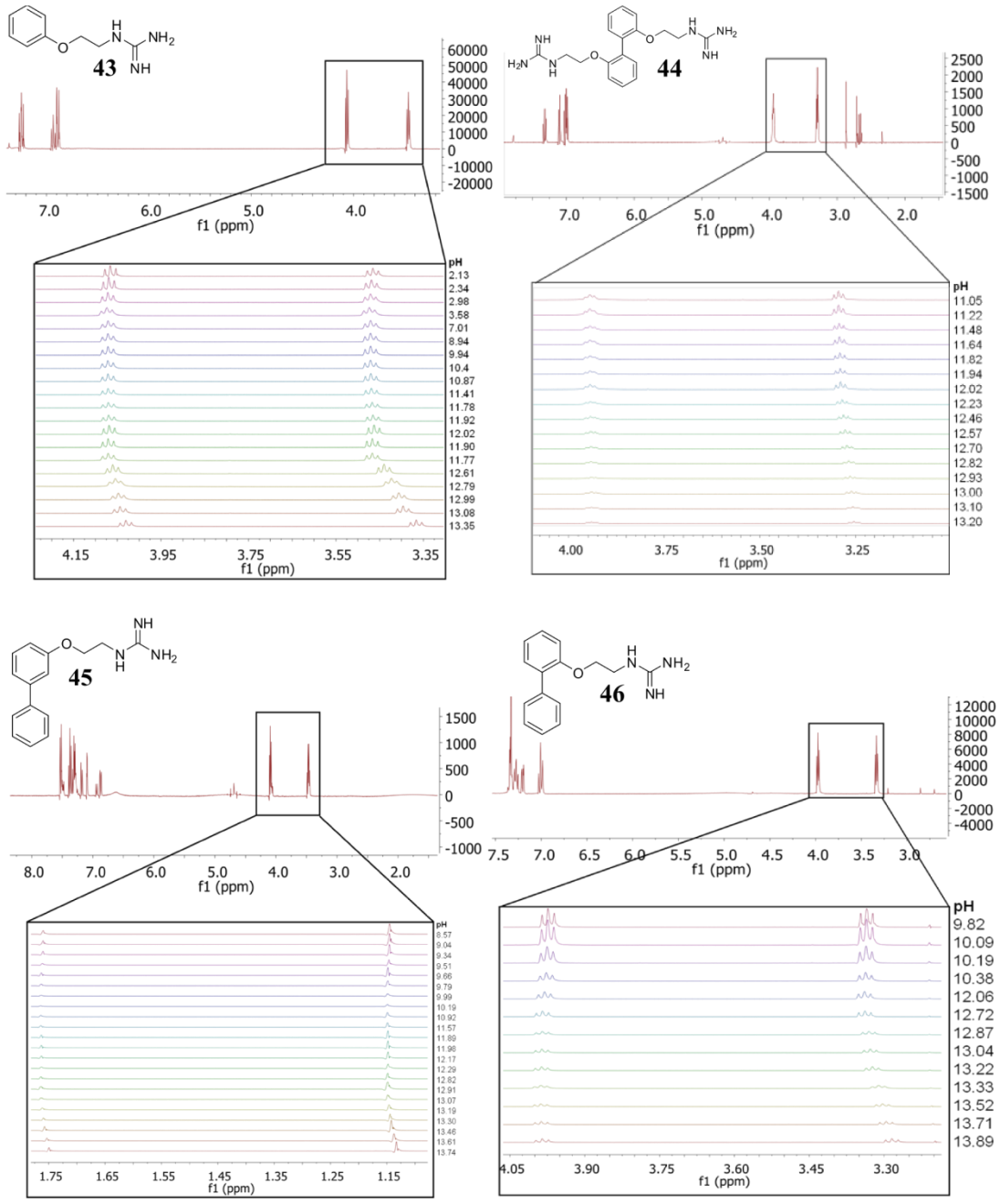


Figure 19. NMR peak shifts due to pH change for compounds 43, 44, 45 and 46. For each compound, the top panel shows the complete 400 MHz ^{1}H-NMR spectra in D_2O. The expanded portion shows the triplets corresponding to the two sets of methylene groups in the compounds. The triplet on left corresponds to the methylene groups closest to the ether oxygen, and the right hand triplet corresponds to the methylene groups adjacent to the guanidine group. As the pH of the solution is increased by adding NaOH, the methylene group closest to the guanidine becomes more shielded as a result of deprotonation of the guanidine group, leading to a greater change in chemical shift than the more distant methylene groups.

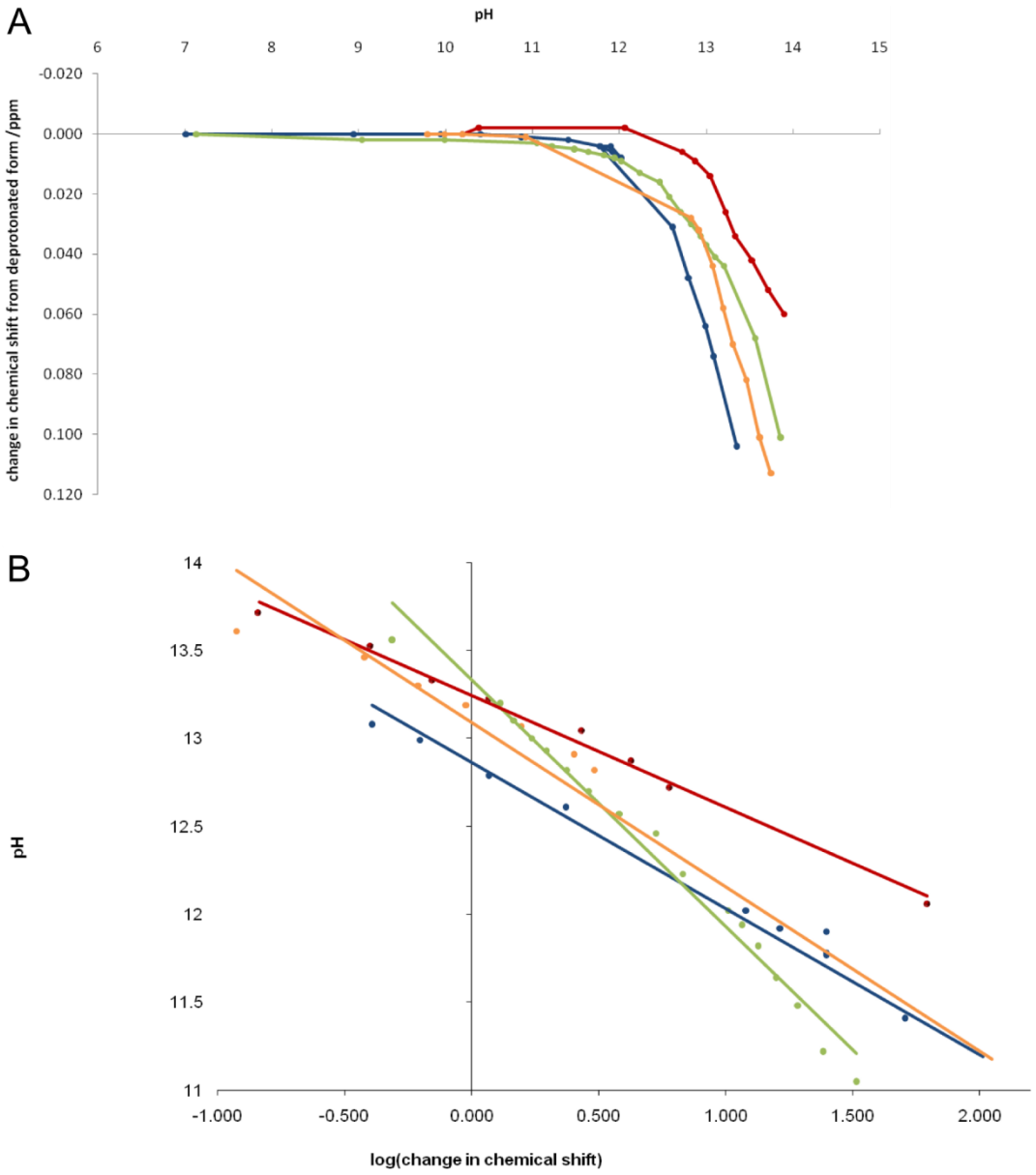
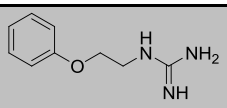
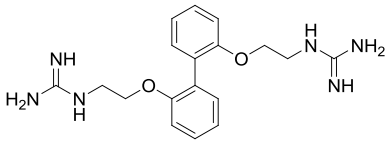
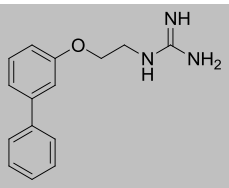
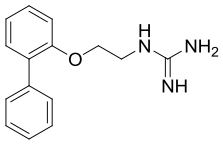


Figure 20. Plots to determine the pK_a values of simple SMoC compounds. **43** (blue), **44** (green), **45** (red) and **46** (orange). A: The change in chemical shifts of the methylene protons adjacent to the guanidines in each compound as a function of pH of the solution. B: The raw data plots were converted to straight lines in order to calculate the pK_a values from the y-intercepts.

Table 9. Calculated pK_a values for the simple SMoC compounds.

Compound	Structure	pK_a
43		12.86
44		13.33
45		13.09
46		13.26

The results of these experiments (Table 9) show that the 2G-SMoC with the bi-phenyl substituted at the ortho position to the biphenyl bond **44** has the highest pK_a (13.33), suggesting that the guanidine groups in this compound benefit from the greatest π-cation stabilisation. The mono-phenyl compound **43** has the lowest pK_a (12.86) and the closest to free arginine (12.48),[146] suggesting very little stabilisation is available to the guanidinium ion. Therefore, π-cation interactions with the opposite phenyl ring are possible, but the length of the alkyl chain is not sufficient for interactions between the guanidines and their adjacent ring. These results also show that substitution position is also important for stabilisation, since the *meta*-substituted compound, **45**, has a lower pK_a (13.09) than the *ortho*-substituted **46** (13.26). This result is in agreement with the predictions made by the molecular dynamics simulations, which predicted that the *meta* compound formed the weaker parallel π-cation interactions rather than the T-shaped interactions seen in the *ortho* compound. Overall, these results suggest that the biphenyl system is able to provide some stabilisation to the positively charged guanidinium ion, probably via

interactions between the cationic guanidine groups and the π-electron cloud of the biphenyl system. The energy of this stabilisation may be calculated by comparing the pK_a values of the mono-phenyl compound **43** with the mono-substituted biphenyl compound **46**. The difference in free energy between the two species may be calculated using the equation

$$\Delta\Delta G^{0}_{\mathrm{p}K_a} = 2.303RT\Delta\mathrm{p}K_a$$

which gives a free energy of stabilisation of -2.3 kJ mol^{-1} for each guanidine group compared to the mono-phenyl compound, for which it was assumed that there is very little π-cation stabilisation, due to its low pK_a. This value is the increase in the free energy of the guanidine group gained by the addition of a phenyl ring to the mono-phenyl compound **43**. The biphenyl compound with the side chain *meta* to the biphenyl bond **45** also showed a small amount of stabilisation, which was predicted by molecular dynamics to be due to the weaker parallel type of π-cation interaction with the opposite ring. Comparing with the mono phenyl compound, the additional stabilisation of this interaction was calculated to be -1.3 kJ mol^{-1}.

3.2.3. 6G-SMoC Synthesis

The reactions in this section were initially carried out by Ann-Sophie Rebstock. The syntheses were repeated by the author at a later date in order to obtain the characterisation data listed in the Experimental Methods chapter and the NMR spectra detailed here.

In order to maximise binding to siRNA, a new SMoC was synthesized with six guanidine groups attached to a terphenyl ring structure. The 6G-SMoC was synthesized (Scheme 18) starting from the alkylated bromoiodocatechol synthesized previously[124] via coupling with the boronic ester, separation of the mono- (**49**) and di-coupling product (**48**) and second coupling of the mono-coupling product. This was followed by deprotection of the Cbz protecting groups and subsequent guanidinylation using the triflate-linked reagent.[147] The Boc protecting groups were deprotected using TFA, and the final product characterised by NMR (Figure 21).

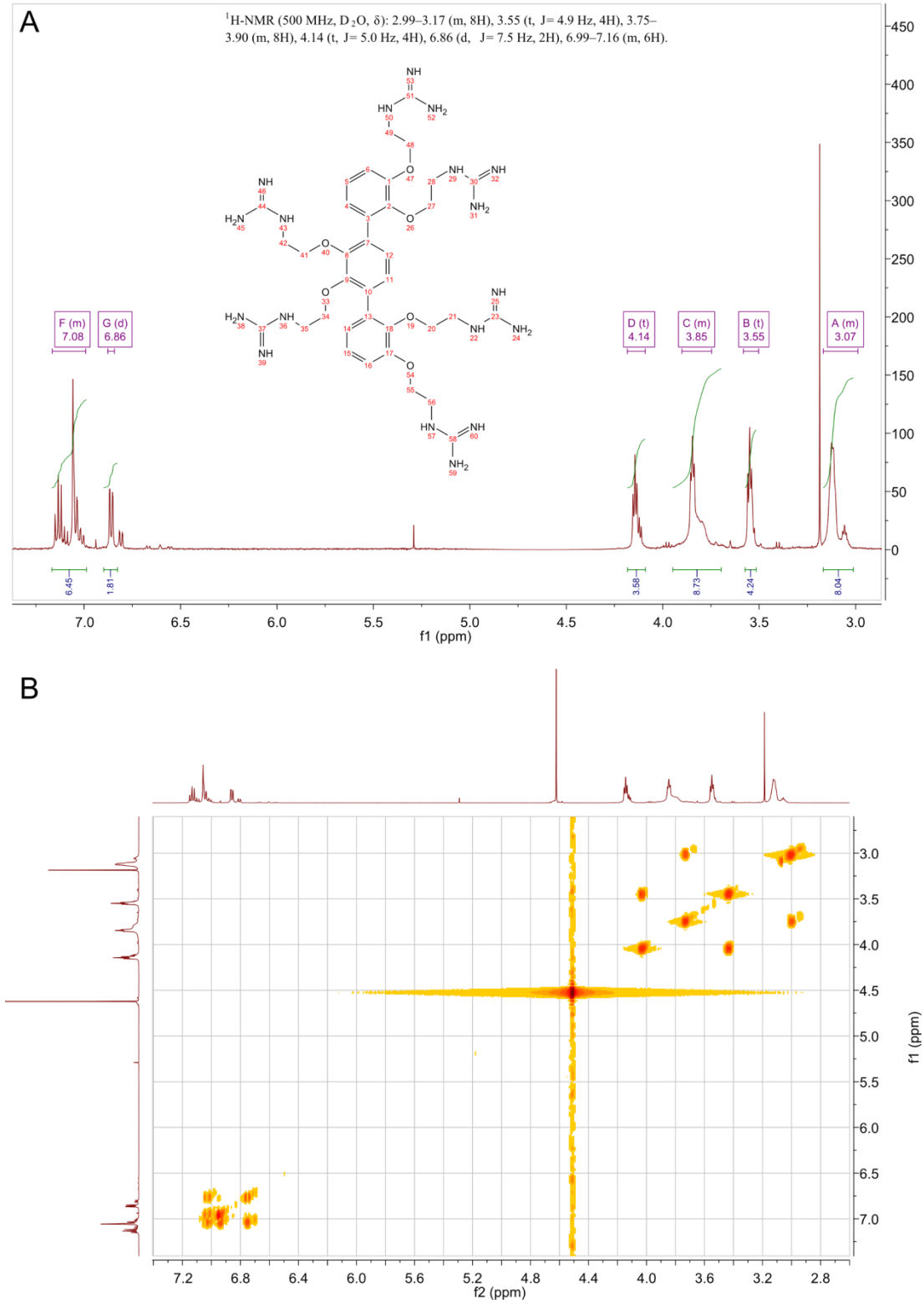
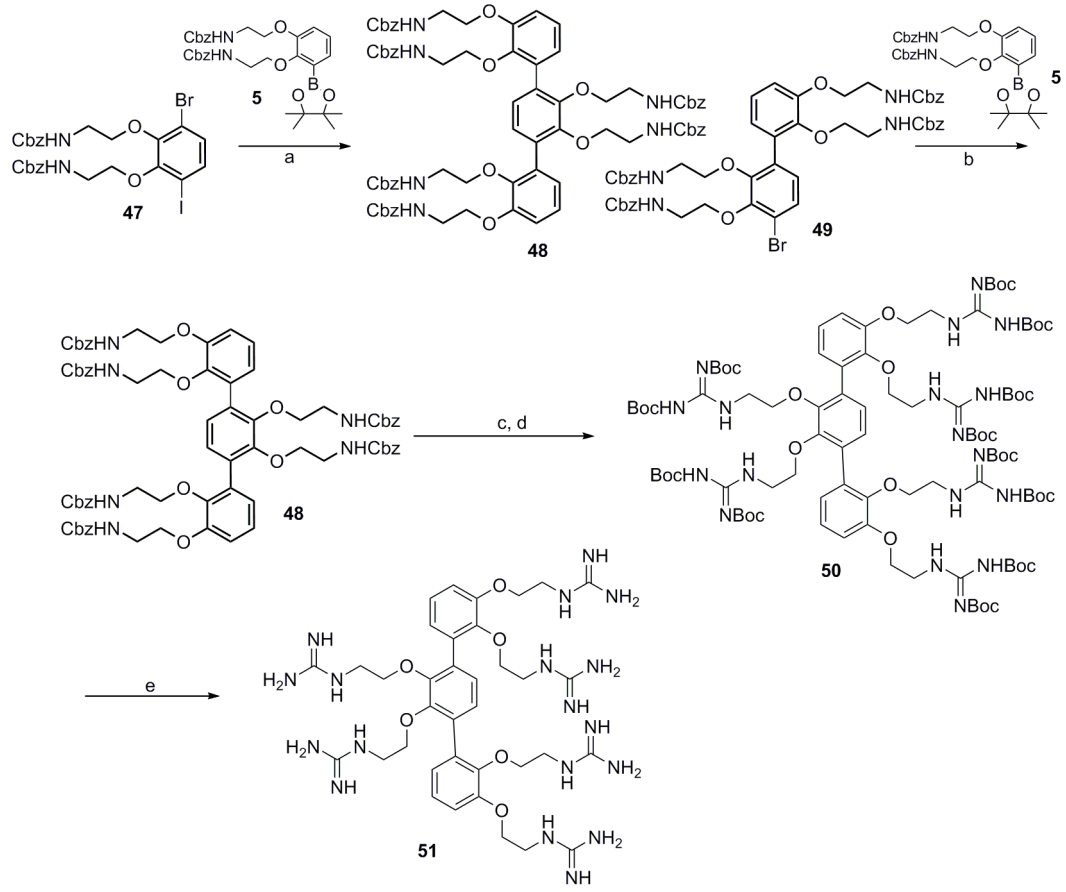


Figure 21. 1D (A) and 2D (B) ^{1}H-NMR spectra of 6G-SMoC. The 2D COSY spectrum shows the splitting pattern of the ethyl protons, allowing full assignment of the spectrum.

The terphenyl structure gives rise to four peaks for the ethyl protons. The COSY spectrum shows coupling between the multiplets at 3.07 and 3.85 ppm, and between the triplets at 3.55 and 4.14 ppm. The 3.07 and 3.85 ppm multiplets integrate to 8 protons each, suggesting that these peaks correspond to the ethyl groups attached to the rings at positions 2, 8, 9 and 18. The triplets at 3.55 and 4.14 ppm integrate to 4 protons each, suggesting that these peaks correspond to the ethyl groups attached to the outside of the molecule at positions 1 and 17.



Scheme 18. Synthesis of 6G-SMoC. Reagents: a) $PdCl_2dppf$, K_3PO_4, toluene, water, 100°C 3h; b) $PdCl_2dppf$, K_3PO_4, toluene, water, 100°C overnight (48% over two steps); c) 30% HBr in acetic acid, DCM, room temperature; d) *N,N*-di-boc-*N'*-trifluoromethanesulfonyl-guanidine, DIEA, DCM, room temperature, overnight (54%); e) TFA/TIPS/H_2O, 4h, room temperature (100%).

3.2.4. ITC

In order to calculate the strength of the electrostatic interactions between SMoC and siRNA, isothermal titration calorimetry (ITC) was used to measure the heat evolved when small aliquots of 6G-SMoC were injected into a solution of a standard 21-bp siRNA for the human housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH). The first experiment carried out involved injecting aliquots of a 10 μM siRNA solution into a 75 μM SMoC solution contained in the calorimeter cell. The resulting data showed that the SMoC was very quickly saturated by the siRNA and therefore the data produced was insufficient to calculate the binding parameters. As an siRNA molecule contains many more charges than a SMoC molecule, it was decided to reverse the experimental setup and place the SMoC in the injector and the siRNA in the calorimeter cell. Injections of 1 mM SMoC into 2 μM siRNA resulted in more data points, however the saturation point was still reached too quickly, meaning that a binding curve could not be fitted (Figure 22A). Finally, by increasing the siRNA concentration to 3.5 μM and decreasing the SMoC concentration to 250 μM, the experiment produced sufficient data to fit a binding curve to calculate the binding parameters (Figure 22B). A binding constant (K) of 87.3 μM was calculated with a binding stoichiometry of 4.4:1 (SMoC:siRNA). Binding was dominated by enthalpic contributions (-1646 Kcal/mol) with a minor or negligible entropic contribution. This is consistent with electrostatic (charge) interactions between the guanidines and the phosphate groups on the siRNA dominating binding.

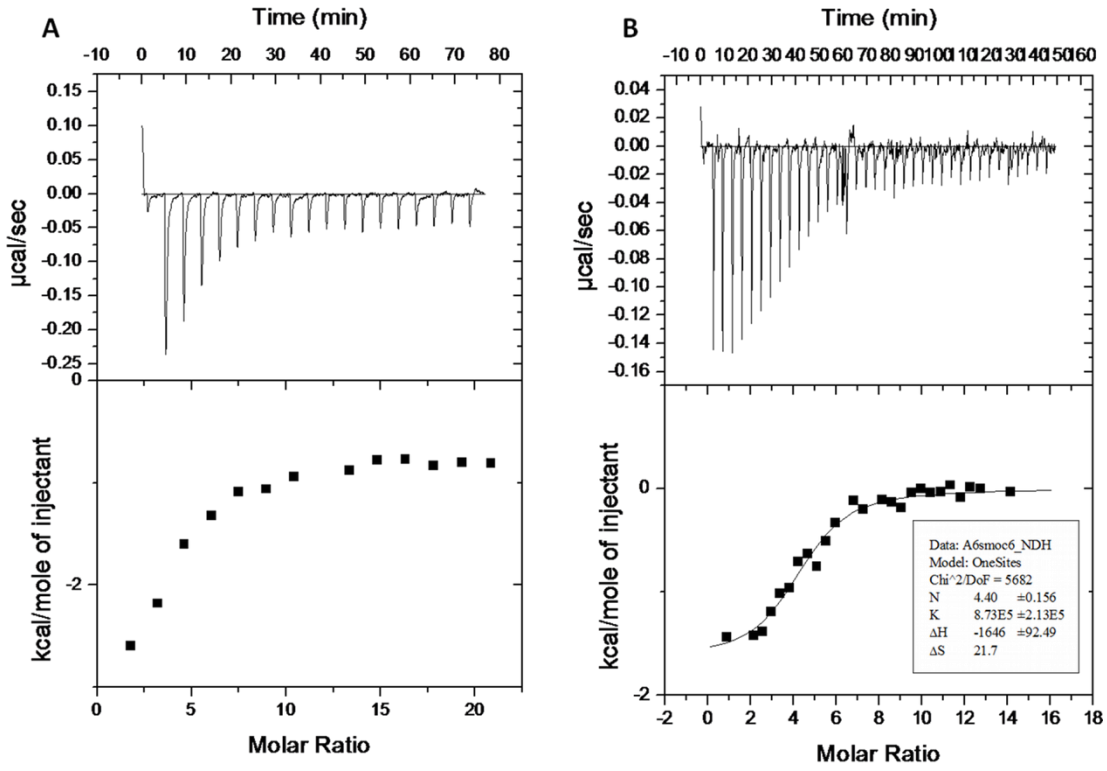


Figure 22. ITC data for titration of 6G-SMoC against human GAPDH siRNA. Above: Each spike represents the heat evolved upon a single 8 µl injection of 6G-SMoC into the siRNA-containing 1.8 ml cell. Below: The Origin software was used to fit the ITC data to a binding curve and calculate the binding parameters. A: 1 mM SMoC injected into 2 µM siRNA. Saturation was reached too quickly, therefore a binding curve could not be fitted. B: 250 µM SMoC injected into 3.5 µM siRNA. Data sufficient for calculation of binding parameters.

3.2.5. Gel Shifts

The interactions of 4G- and 6G-SMoC compounds with siRNA were analysed by gel electrophoresis of SMoC/siRNA complexes at different molar ratios. Tetra- and octa-arginine peptides were also tested in order to establish the relative binding affinity of SMoC compared to these established cell delivery peptides. All SMoCs and peptides were able to fully shield the siRNA negative charges at ratios less than 20:1 (Figure 23). All the gels showed a slightly shifted siRNA band at low SMoC/siRNA ratios due to partial charge neutralisation, but this band was absent for higher ratios, suggesting complete complexation of the siRNA and charge neutralisation, preventing EtBr staining. Since the EtBr intensity of the observable

band on the gel is proportional to the free siRNA concentration we utilised this value to provide a convenient graphical representation of the relative affinities and to calculate the approximate binding constants of the analogues.

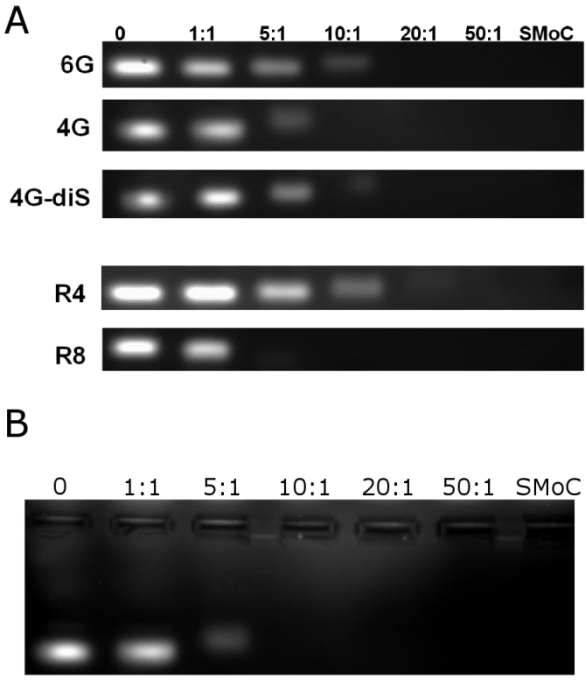


Figure 23. Gel shift assay to determine binding affinities of SMoCs and oligoarginine peptides to GAPDH siRNA. A: GAPDH siRNA was mixed with different SMoC compounds (6G- 4G- or 4G-disulfide) or a polyarginine peptide (R4 or R8) at the ratios shown (SMoC:siRNA, 0 = siRNA only, SMoC = SMoC/peptide only) and the mixtures incubated at room temperature for 30 min. The samples were diluted with a running dye containing EtBr and run on a 1% agarose gel. The gels were visualised under UV illumination and the single band corresponding to the GAPDH siRNA for each sample is shown. B: Expanded gel of 4G-SMoC showing that the siRNA band is slightly shifted at 5:1 SMoC:siRNA, but the band is absent at higher ratios due to complete complex formation.

The EtBr fluorescence of each band was quantified and the data plotted as dose response curves in order to determine the binding affinity of each compound to siRNA (Figure 24). 6G-SMoC had a slightly lower binding affinity than 4G-SMoC (EC_{50} of 0.11 mM compared to 0.07 mM) and 4G-SMoC had a higher affinity than R4 (EC_{50} of 0.07 mM compared to 0.12 mM). R8 had a higher affinity (0.03 mM)

than any of the SMoCs. The disulfide variant of 4G-SMoC has the same binding affinity as the monomer form, probably due to reduction of the disulfide bond in the gel running buffer due to the presence of MOPS, a mild reducing agent. 4G-SMoC binds with a higher affinity than the polyarginine peptide R4, indicating an ability to complex to siRNA more efficiently or to shield the negative charges more effectively. This may be attributed to the increased pK_a of the SMoC guanidine groups as a result of the pi-cation effect described previously. The large Hill coefficients calculated indicate the stoichiometry is greater than 1:1, which is in agreement with the ITC data.

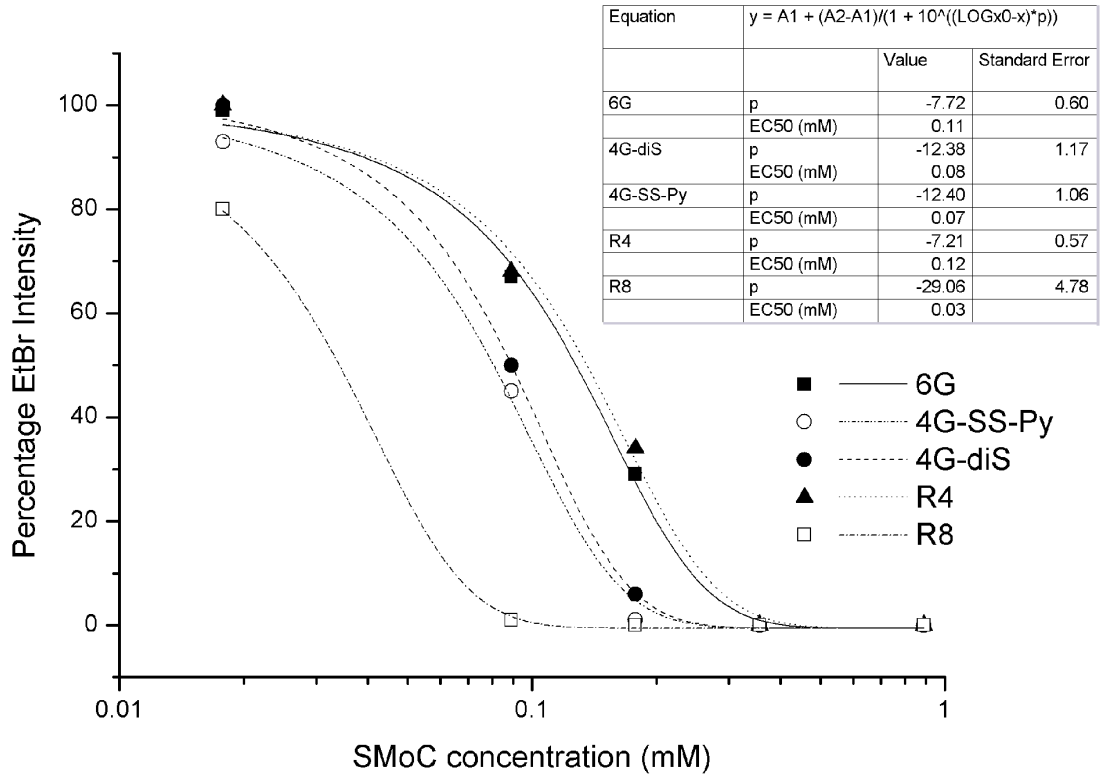


Figure 24. Dose response curves for the binding of SMoCs and oligoarginine peptides to GAPDH siRNA. The EtBr bands from Figure 23 were quantified and the band intensity plotted to produce binding curves for each of the compounds with siRNA. The EC_{50} values and Hill coefficients (p) of each compound are shown, giving a relative indication of the binding affinity of each compound for siRNA.

3.2.6. siRNA transfection

In order to determine if SMoCs are capable of transporting siRNA into cells resulting in a knockdown of the target gene, IMR-90 primary human fibroblasts and Human Embryonic Kidney (HEK)-293 cells were transfected with SMoC or Lipofectamine complexed with either a non-targeting control siRNA, or an siRNA targeted at *cdc7*. This gene is a cell cycle kinase which promotes entry into the S phase of the cell cycle in which DNA synthesis takes place and has been identified as a possible therapeutic target in epithelial ovarian carcinoma.[148] Lipofectamine was used according to the manufacturer's instructions, and SMoC was used in a large molar excess (1:5000) in order to ensure complete saturation of the siRNA by SMoC. After 72h, the cells were harvested, the RNA purified and the *cdc7* mRNA measured using qRT-PCR (Figure 25A). Both the siRNA sequences, and the PCR primers for c*dc7* and GAPDH have previously been validated in human cancer cells.[149] In order to confirm the quality of the PCR amplification, a melting curve analysis was carried out. This involves slow heating of the PCR samples following the amplification cycles and measurement of the changes in fluorescence. When the DNA in the sample reaches its melting point, the strands separate and SYBR is released, resulting in a drop in fluorescence. Each DNA sequence has a unique melting point depending on its constituent bases, resulting in a characteristic peak on the melting curve. By plotting the change in fluorescence against temperature, it may be established how many amplification products are present. In this case, two peaks were observed across all the samples corresponding to c*dc7* and GAPDH DNA which were amplified in the PCR reaction **(**Figure 25B). For each compound tested, the *cdc7* mRNA level was calibrated against a set of untreated cells (UT). The cells transfected with non-targeting siRNA (CO) were used to take into account any effect that the compound itself has on mRNA expression. The total knockdown was calculated as the difference between the CO cells and the cells transfected with *cdc7* siRNA. In both HEK-293 and IMR-90 cells, the difference between CO and *cdc7* transfected cells using 6G-SMoC was insignificant ($p=0.22$ for HEK-293, $p=0.41$ for IMR-90). 4G-SMoC did result in a significant difference between CO and *cdc7* transfected cells, and produced a knockdown of 19.2% in HEK-293 and 36.75% in IMR-90 cells. These results were similar to the knockdown achieved using the

commonly used transfection reagent Lipofectamine, which produced knockdowns of 36.6% (HEK-293) and 28.1% (IMR-90).

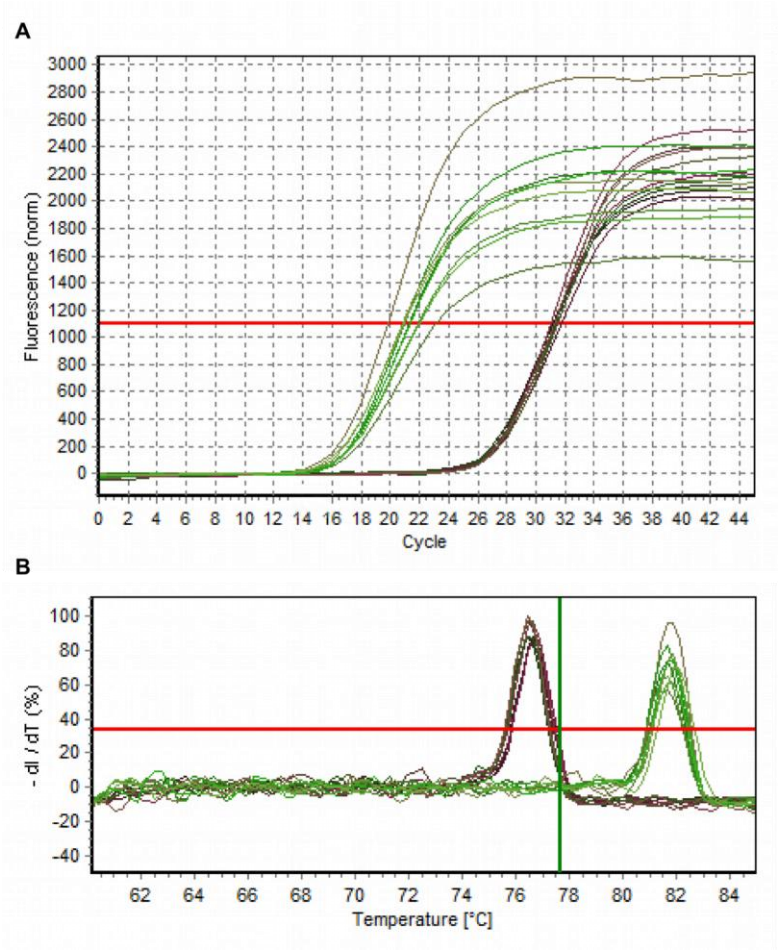


Figure 25. Raw PCR data for RNA extracted from IMR-90 cells either untreated, or treated with 4G-SMoC-SSPy complexed with either control or *cdc7* siRNA. A: Fluorescence as a result of SYBR binding to DNA increases exponentially as the number of amplification cycles increases until exhaustion of the primers. The two groups of samples represent incubation with GAPDH (left) or Cdc7 (right) primers. The red line indicates the threshold at which the C_t value is measured; B: Melting curve analysis of the PCR product in each sample. The peak on the left corresponds to Cdc7 DNA and the peak on the right corresponds to GAPDH DNA. Each sample was run in triplicate.

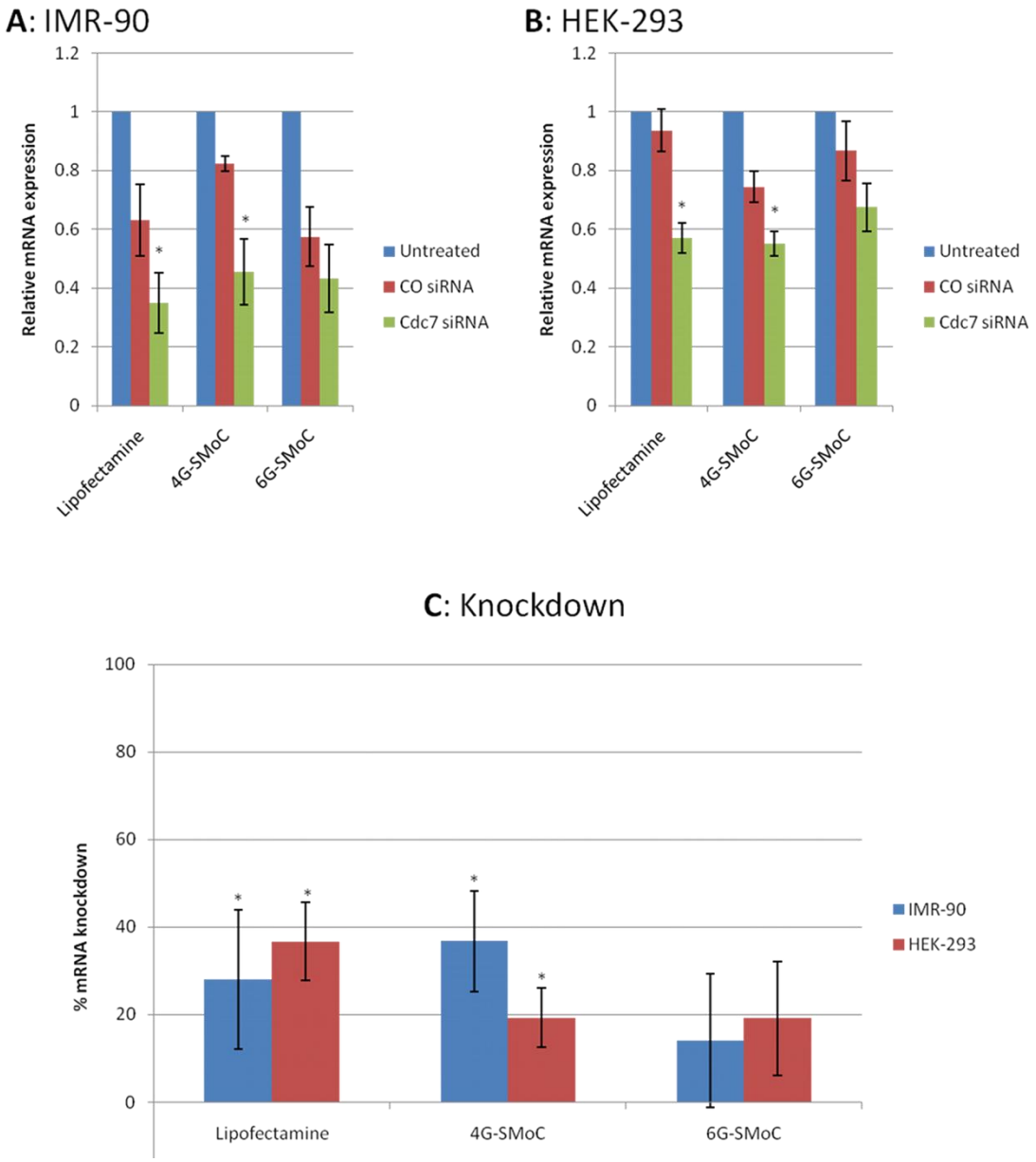


Figure 26. mRNA knockdown by SMoC-siRNA or Lipofectamine-siRNA complexes. IMR-90 (**A**) and HEK-293 (**B**) cells were transfected with a negative control (CO) or *cdc7* siRNA using Lipofectamine, 4G-SMoC and 6G-SMoC. The RNA was purified and the amount of *cdc7* mRNA was quantified by qRT-PCR using untreated (UT) cells as a calibrator. Knockdown (**C**) was calculated as the difference in *cdc7* expression between cells transfected with the negative control siRNA and cells transfected with *cdc7* siRNA. Experiments were conducted in triplicate. *, P<0.05 compared to CO siRNA.

3.3. Discussion

Since SMoCs have been shown to deliver active proteins into cells, it was decided to apply this new class of transporters to the delivery of siRNA, which has many therapeutic and commercial applications. Since SMoCs are highly positively charged, the most obvious way of linking them to siRNA is via the formation of electrostatic complexes, which eliminates the requirement of a further synthetic step, and can be applied to any unmodified siRNA sequence. It was predicted that SMoC-siRNA complexes may be used to deliver siRNA to the cytoplasm, resulting in mRNA knockdown.

Molecular dynamics simulations followed by measuring the pK_a values of simple SMoC compounds by NMR spectroscopy have confirmed the hypothesis that SMoC compounds contain an intramolecular π-cation interaction suggesting that the guanidine groups in SMoCs are stabilised by π-cation interactions with the electron cloud of the biphenyl rings (Figure 27). The additional charge stabilisation may lead to tighter anion binding, which is significant both for the formation of SMoC-siRNA complexes as well as binding to cell surface proteoglycans, such as HS and CS, and phospholipids, which may act as a precursor to internalisation. In addition, π-cation interactions may enhance membrane insertion by shielding the cations from the water phase as suggested for penetratin.[143]

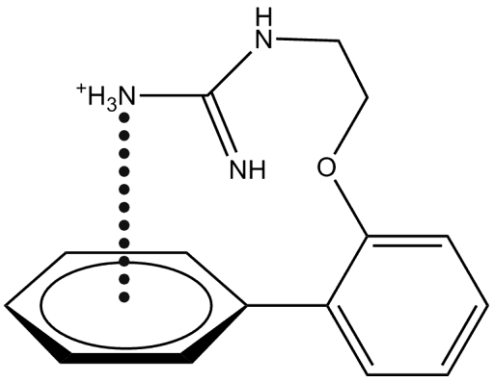


Figure 27. Illustration of the π-cation effect for a biphenyl SMoC molecule.

The new guanidinylation reagent identified previously was particularly useful in the synthesis of the simple SMoCs used in this study. The proton NMR spectrum of Boc-protected biphenyl compound **40** was observed to have a unexpectedly high chemical shift values for all peaks which were higher than for all other SMoC compounds. This is thought to be due to incorrect calibration to the internal TMS standard when measuring the spectrum, since the spectrum of the deprotected compound **44** is as expected.

As hypothesized, various SMoC compounds were capable of forming electrostatic interactions with siRNA, forming netural complexes. Using ITC, the binding affinity of 6G-SMoC to siRNA was accurately measured as 87.3 μM, and the relative affinities of other SMoC compounds, as well as tetra- and octaarginine peptides, were measured using a gel shift assay. The binding affinities of the SMoCs and peptides to siRNA does not seem to be related to the number of guanidine groups, since 6G-SMoC has a lower affinity than 4G-SMoC-SSPy. The higher binding affinity of 4G-SMoC-SSPy than R4 supports the π-cation hypothesis that the guanidine groups of the SMoC compounds are stabilised, and hence have a higher anion affinity. The weaker binding of 6G-SMoC than 4G-SMoC-SSPy may be explained by the presence of the linker group on 4G-SMoC-SSPy which may act as an electron donating group via induction from the methylene group. Ideally, this experiment should be carried out using SMoCs without linker groups in order to further establish the connection between number of guanidine groups and binding affinity to siRNA. However, this experiment confirms that SMoC-siRNA complexes may be formed in solution in a similar way to CPPs and that that 4G-SMoC-SSPy has the highest affinity for siRNA, and therefore shows the greatest potential for siRNA transfection.

Transfection using SMoCs was carried out in two cell lines; IMR-90 cells, a human primary cell line, and HEK-293 cells, an immortalised cell line. As predicted, SMoC-siRNA non-covalent complexes were capable of delivering functional siRNA to the cytoplasm, resulting in signifcant levels of mRNA knockdown. This is a

significant new development for the SMoC compounds, as previously they have only been shown to deliver covalently conjugated proteins and dyes. 4G-SMoC-SSPy achieved a comparable level of knockdown to Lipofectamine, suggesting that this compound has potential as an alternative *in vitro* reagent. However, having established this potential, further uptake studies should be carried out in order to compare 4G-SMoC-SSPy against other amphipathic peptide vectors such as penetratin. If the observed π-cation interaction is involved in the internalisation mechanism as described for penetratin, it may be expected that SMoCs will show a similar uptake efficiency to the peptide. 6G-SMoC did not achieve a significant gene knockdown, suggesting that affinity for siRNA is related to transfection ability. The knockdowns recorded for Lipofectamine and 4G-SMoC-SSPy are significant compared to untreated cells, however the knockdown achieved for Lipofectamine is lower than in similar studies reported in the literature,[105, 111-112] suggesting that the experimental procedure used is not fully optimised. Further optimisation of the methods may therefore result in more significant silencing using the SMoC compounds. Time restrictions prevented a full optimisation study of the experimental conditions, including the effect of cell culture media, siRNA and SMoC concentrations, and the ratio of siRNA to SMoC. In particular, it should be tested whether using SMoCs at the stoichiometric ratio indicated by the gel shift assays, rather than in large excess, may improve uptake. This may be significant, since a large excess of free SMoC in the transfection mixture may compete for binding to cell surface GAGs with the siRNA-SMoC complexes. Since the guanidine groups of the complexed SMoC molecules are engaged in electrostatic interactions with siRNA, the uncomplexed SMoC may have a higher affinity for GAGs, and therefore may block binding of the siRNA complexes, limiting siRNA internalisation.

An interesting result from this study was that 4G-SMoC-SSPy, originally developed as a protein internalisation reagent due to its ability to form disulfide linkages with cysteine residues was more successful at promoting siRNA transfection than 6G-SMoC, which possesses more guanidine groups. It has been shown that a nonaarginine (R9) peptide with terminal cysteine residues forms high molecular

weight (~96 kD) complexes which are able to condense DNA and promote significantly higher gene uptake than momomeric R9.[150] The formation of disulfide linked dimers could explain the higher mRNA knockdown of 4G-SMoC-SSPy than 6G-SMoC in complex with siRNA. The purified disulfide form of 4G-SMoC **34** has not yet been tested in transfection assays, but it is hypothesized that it will show similar or higher uptake than 4G-SMoC-SSPy. Furthermore, it has been suggested that CPPs containing intra- or intermolecular disulfide bonds may be cross-linked by cell surface thiol proteins, resulting in a higher uptake efficiency.[151] This hypothesis may also apply to disulfide-containing SMoCs, which has implications for the design of future compounds.

CHAPTER 4

Designing a new SMoC for siRNA delivery

4. Designing a new SMoC for siRNA delivery

4.1. Introduction

In the previous chapter, it was shown that 4G-SMoC is capable of delivering siRNA across cell membranes to achieve a significant gene silencing effect via the RNAi mechanism. The small mRNA knockdown achieved by SMoC-siRNA complexes implies that the siRNA is not available in the cytoplasm in high enough levels to completely silence the gene, suggesting that the complexes are not sufficiently taken up by the cells, or are trapped in endosomes preventing the siRNA from reaching its target. Studies on CPPs have shown that binding to cell surface GAGs leads to clustering of the attached proteoglycans resulting in reorganisation of the actin skeleton which is an important precursor for internalisation by endocytosis.[31] It is therefore hypothesized that in order to achieve a high level of uptake, SMoCs must have the ability to optimally bind these cell surface anions as well as siRNA. Two of the most prevalent cell surface GAGs are HS, which is a component of proteoglycans at the cell surface including transmembrane syndecans and GPI-anchored glypicans, and CS which is the most prevalent GAG and has both structural and regulatory roles in the extracellular matrix.[152] Therefore, it is predicted that by analysing the crystal structures of these GAGs, a new SMoC compound may be designed which retains the siRNA-complexing properties but has improved uptake efficiency than the previosuly described SMoCs.

As discussed previously (1.3.2.5), one of the features of a multi-component siRNA delivery system is a targeting sequence in order to target the siRNA to a particular location *in vivo*,[118] for example a particular cell type via a cell surface receptor. It is hypothesised that in addition to providing selectivity, binding to cell surface receptors may also increase uptake efficiency by increasing the concentration of the siRNA delivery vehicle at the cell surface as observed by nanaoparticles bearing cell-specific targeting ligands.[153] RGD peptides containing an arginine-glycine-aspartate motif have been found to be potent ligands for cell surface integrins, which are present on all cells and mediate binding to the extracellular matrix. In addition, integrins carry out intracellular signalling roles, with overexpression of certain

integrins acting as indicators of diseases such as cancer, thrombosis and autoimmune diseases.[154] Specificity for specific integrins is determined by the secondary structure of the RGD motif, and the residues flanking it. Various commercially available RGD peptides have been produced that target different integrins in order to target specific disease states. The cyclic RGD peptide cyclo(Arg-Gly-Asp-D-Phe-Lys) (cRGDfK), for example, contains a γ-turn at the Gly residue, and is specific for the $\alpha_V\beta_3$ and $\alpha_V\beta_5$ integrins which are overexpressed in many tumours.[155] This peptide was chosen for use in this study as its target integrins are overexpressed in many common cancer cell lines. It is hypothesized that this targeting peptide may be conjugated to the new SMoC compound using a carboxylic acid linker group via a peptide coupling reaction. Therefore, the design of the new SMoC should include this functionality. Since the peptide is large in comparison to the SMoC compound, it is predicted that a PEG linker is required to separate the SMoC from the peptide in order that the guanidine groups are not sterically hindered by the peptide, blocking interactions with the siRNA. The cyclic peptide with a PEG linker attached to the lysine residue is commercially available (Figure 28).

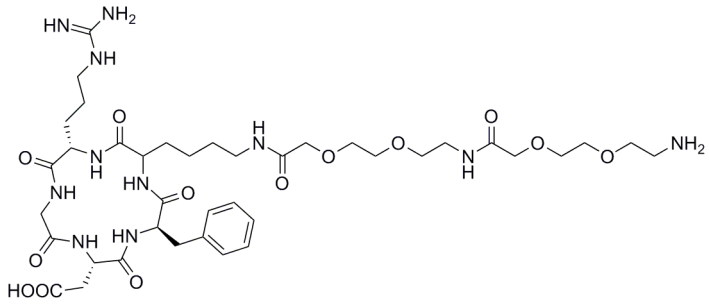


Figure 28. cRGDfK peptide with a $[PEG_2]_2$ linker attached to the lysine residue. The terminal amine allows peptide coupling to acids.

Therefore the aims of this chapter are to analyse the crystal structure of common GAGs to determine if SMoCs may be optimised for cell surface binding; to design a synthesis of this SMoC which includes an acid linker group for conjugation to the RGD peptide; to test whether this SMoC retains its ability to complex siRNA; and to test whether the new compound displays increased cell uptake capabilities by

measuring mRNA knockdown in cultured cells treated with SMoC-siRNA complexes.

4.2. Results

4.2.1. Design of a new SMoC analogue

To optimise SMoCs for binding to these GAGs, crystal structures were obtained from the Protein Data Bank (PDB) in order to measure the distance between sulfate residues. Since no crystal structure of HS was available, the structure of heparin was obtained instead (PDB ID: 1HPN).[156] Heparin is a highly sulfated GAG whose structure is closely related to heparan sulfate. Both GAGs consist of a repeating disaccharide unit; in HS, the most common unit is glucuronic acid linked to *N*-acetylglucosamine (GlcA-GlcNAc), which makes up ~50% of the structure, whereas heparin is ~75-85% comprised of 2-*O*-sulfo-α-L-iduronic acid linked to 2-deoxy-2-sulfamido-α-D-glucopyranosyl-6-*O*-sulfate (IdoA(2S)-GlcNS(6S)). The GlcA-GlcNAc repeats of HS undergo modifications to 2-deoxy-2-sulfamido-α-D-glucosamine (GlcNS), iduronic acid (IdoA), IdoA(2S) and GlcNS(2S).[157] The sulfate residues of HS are mainly located in sulfated S-domains linked by non-sulfated domains, and so the structure of heparin may provide a reasonable approximation of the S-domains of HS.[38] Examination of the crystal structure of heparin revealed that the sulfate residues were positioned in 'clusters' of three on alternating sides of the sugar chain. The distance between residues within each cluster is approximately 6 Å, whereas the distance between clusters on the same side of the chain is approximately 17.5 Å (Figure 29).

The crystal structure of CS was also obtained from the PDB (PDB ID: 2C4S).[158] CS consists of alternating *N*-acetylgalactosamine (GalNAc) and glucuronic acid (GlcA) residues, some of which undergo sulfation by sulfotransferases, particularly at the 4 and 6 positions of GalNAc and 2 position of GlcA. For the crystal structure obtained, the distance between sulfate groups was measured as 12.7 Å, with sulfates positioned on alternating sides of the sugar chain. Since the SMoC guanidine groups may not be able to access sugars on opposite sides of the chain, the distance between

sulfates on the same side was also measured and found to be approximately 20.7 Å (Figure 30).

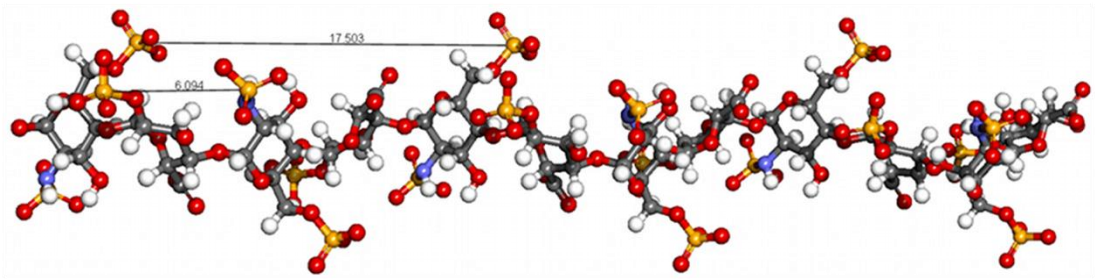


Figure 29. Examination of the crystal structure of heparin. Clusters of three sulfate residues are observed, occurring on alternating sides of the sugar chain. The distances between residues in the same cluster, and between residues of adjacent clusters on the same side of the chain are shown in Å.

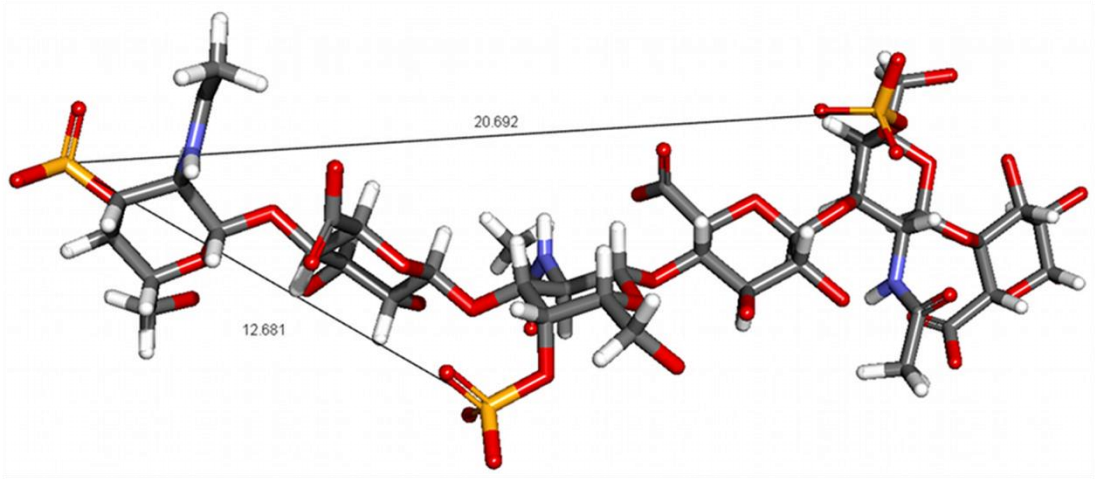


Figure 30. Chondroitin sulfate crystal structure obtained from the PDB. The distances between sulfate ions are marked in Å.

The structure of 4G-SMoC-SSPy, the most successful compound in terms of siRNA transfection, was then analysed to determine its ability to bind the GAGs. An energy minimised structure was produced using the molecular dynamics software YASARA

in order to give a realistic conformation, and the distance between guanidine groups measured. The distance between guanidines attached to the same ring was measured at ~9 Å and the maximum distance between the groups was found to be ~17 Å (Figure 31). This maximum distance is sufficient for binding to heparin sulfates in the same cluster, as well as binding to CS sulfates on opposite sides of the sugar chain. However, in order to optimise binding to these GAGs, the distance between guanidine residues should be slightly increased in order that it is possible to bridge the gap between heparin sulfate clusters, since HS is less sulfated than heparin, and the sulfate groups may be more sparsely arranged. Furthermore, macropinocytosis is initiated by clustering of GAG chains attached to different proteoglycan molecules.[38] Therefore, the increased spacing between SMoC guanidine groups may allow for binding of sulfates on different GAG chains, thereby increasing proteoglycan clustering.

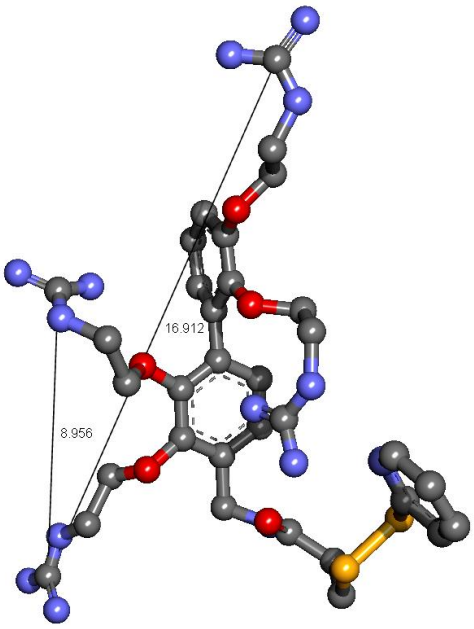


Figure 31. The energy minimised structure of 4G-SMoC-SSPy. The distances between guanidine groups are indicated in Å.

In order to extend the distance between guanidine groups in the SMoC structure, an additional phenyl ring may be added to produce a dendritic structure. The ring

should be added at the same location as the biphenyl bond in order to preserve the π-cation stabilisation of the guanidine charges. The resulting 4G-SMoC-dendrimer (4G-SMoC-Dr) structure was energy minimised in YASARA, and the distance between guanidines measured at ~20 Å (Figure 32), which is sufficient for binding to adjacent sulfate clusters in heparin, and therefore may also bind more effectively to HS at the cell surface. The additional distance between guanidines may also allow binding of additional sulfate groups on CS, as well as increasing the possibility of promoting clustering of GAG chains.

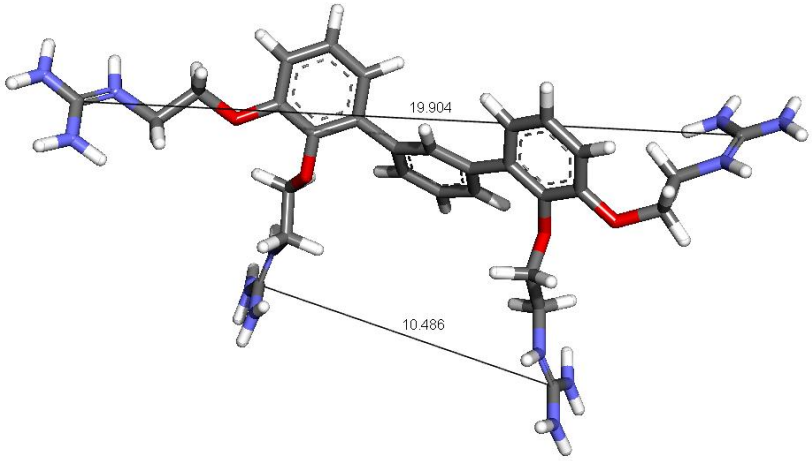
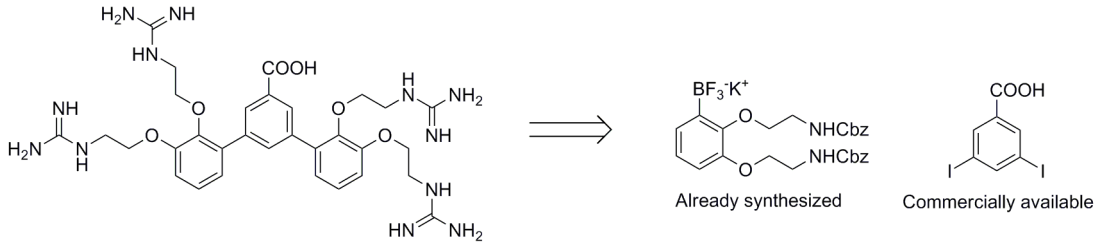


Figure 32. The energy minimised terphenyl structure of 4G-SMoC-Dr. The distances between guanidine groups are indicated in Å.

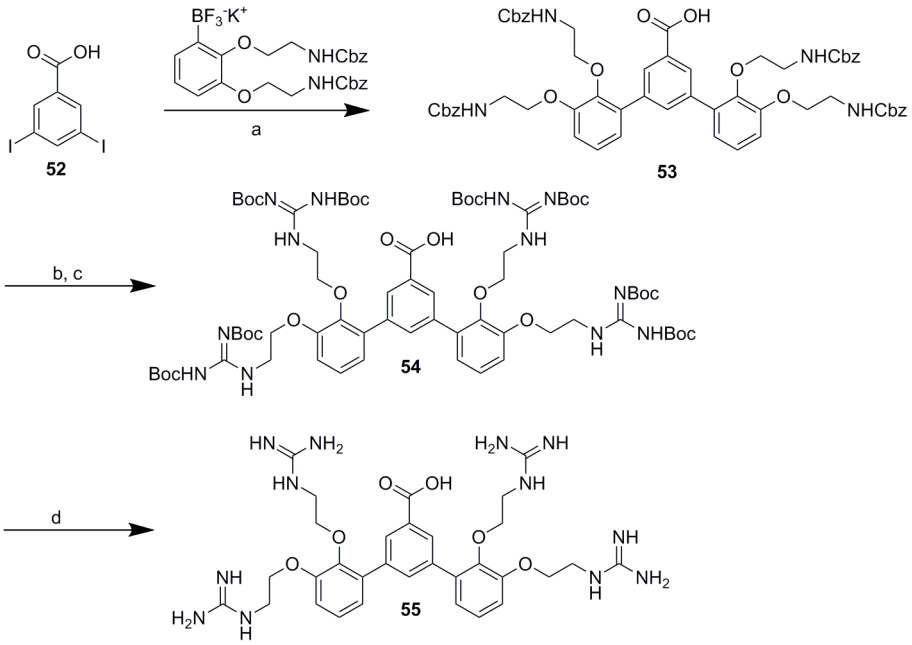
4.2.2. 4G-SMoC-Dr synthesis

Having determined that the new SMoC analogue should contain a terphenyl structure, the synthesis of the compound was planned (Scheme 19). It was decided to include an acid linker group in 4G-SMoC-Dr in order that the cyclic RGA peptide could be linked via amide coupling to the free amine. The synthesis would make use of the trifluoroborate salt **18** described previously, which had been synthesized in bulk, in order to introduce the amine functionality for adding guanidine groups. The additional phenyl ring was added via 3.5-diiodobenzoic acid, a commercially available compound that also contains an acid for use as a linker group.



Scheme 19. Retrosynthetic analysis of dendritic SMoC.

The new SMoC was synthesized via the reactions described previously in three steps (Scheme 20). The terphenyl structure was created using a Suzuki-Miyaura palladium bicoupling reaction under microwave irradiation at 80°C with the trifluoroborate salt **18** and triethylamine, producing the bi-coupled product in 80% yield following isolation by flash column chromatography. Since hydrolysis of the ester in the previous synthesis of a 4G-SMoC with a *meta* linker (Section 2.2.2) proved difficult, it was decided to attempt the guanidinylation in the presence of the acid. The Boc-protected guanidines were added using the newly identified pyrazole guanidinylation reagent with DIEA in 63% yield after purification, demonstrating that the acid does not preferentially react with the guanidinylation reagent. The guanidines were deprotected with TFA and TIPS to yield the final product in quantitative yield which was characterised by NMR spectroscopy (Figure 33). The spectrum shows four triplets corresponding to the 16 ethyl protons on the side chains. This demonstrates the symmetry in the molecule, since the protons in the side chains on either side of the linking phenyl ring are equivalent. The four triplets correspond to the methylene protons adjacent to either the ether oxygen and the guanidine groups for the side chains *ortho* or *meta* to the biphenyl bond. The splitting patterns of the peaks corresponding to the aromatic protons allow for complete assignment of the spectrum.



Scheme 20. Synthesis of 4G-SMoC-dendrimer. Reagents: a) **18**, $PdCl_2dppf.CH_2Cl_2$, Et_3N, iPrOH/H_2O, MW 80°C (80%), 10 min; b) 30% HBr in acetic acid, DCM, rt; c) *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine, DIEA, DCM, room temperature, 17h (63% over 2 steps); d) TFA/TIPS/H_2O, rt, 4 h (100%).

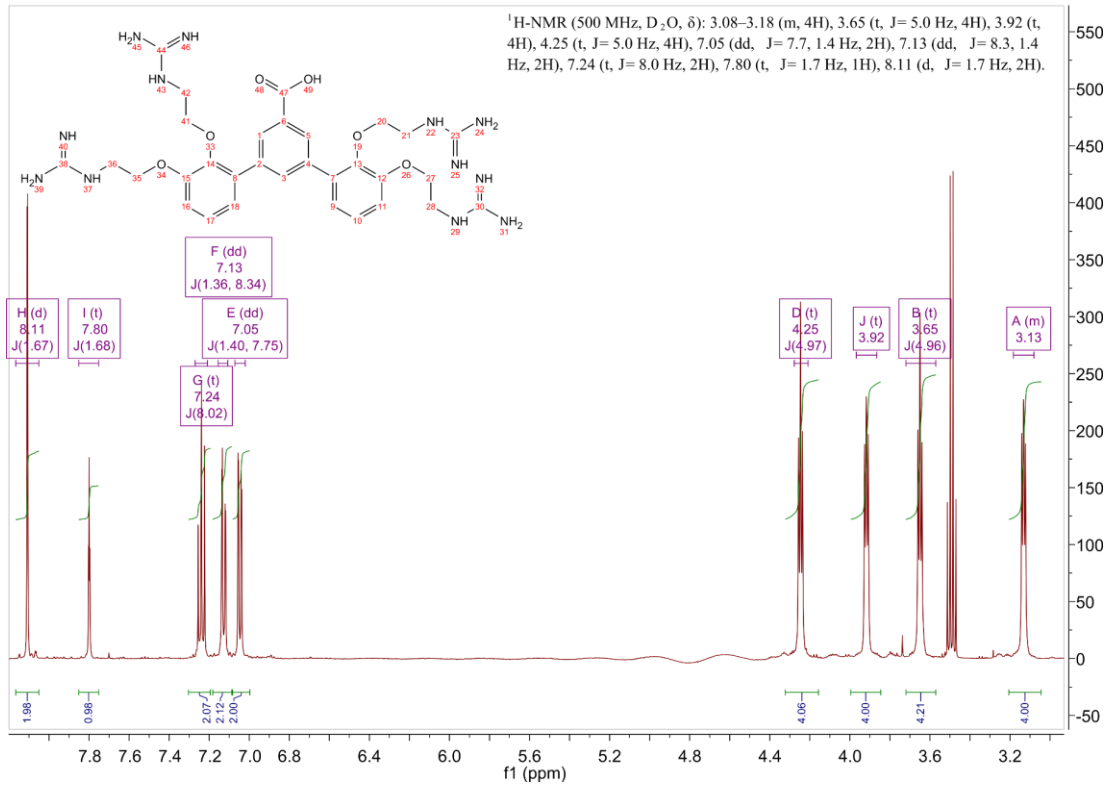
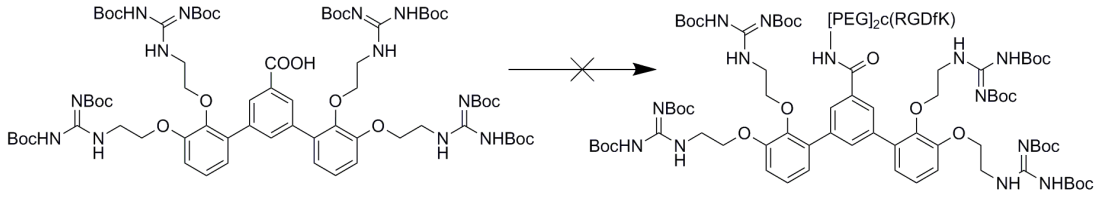


Figure 33. ^{1}H-NMR spectrum of 4G-SMoC-Dr. The triplets at 3.13 and 3.65 ppm correspond to the methylene protons adjacent to the guanidine groups either *meta* or *ortho* to the biphenyl bond respectively, whereas the triplets at 3.92 and 4.25 ppm correspond to the methylene protons adjacent to the ether oxygen in the side chains *meta* and *ortho* to the biphenyl bond. The splitting pattern suggests that the peak at 7.05 ppm corresponds to the protons at positions 11 and 16, the peak at 7.13 ppm to the protons at 9 and 18 and the triplet at 7.24 ppm to the protons at 10 and 17. The three protons on the central ring correspond to the peaks at 7.80 ppm (position 3) and 8.11 ppm (positions 1 and 5).

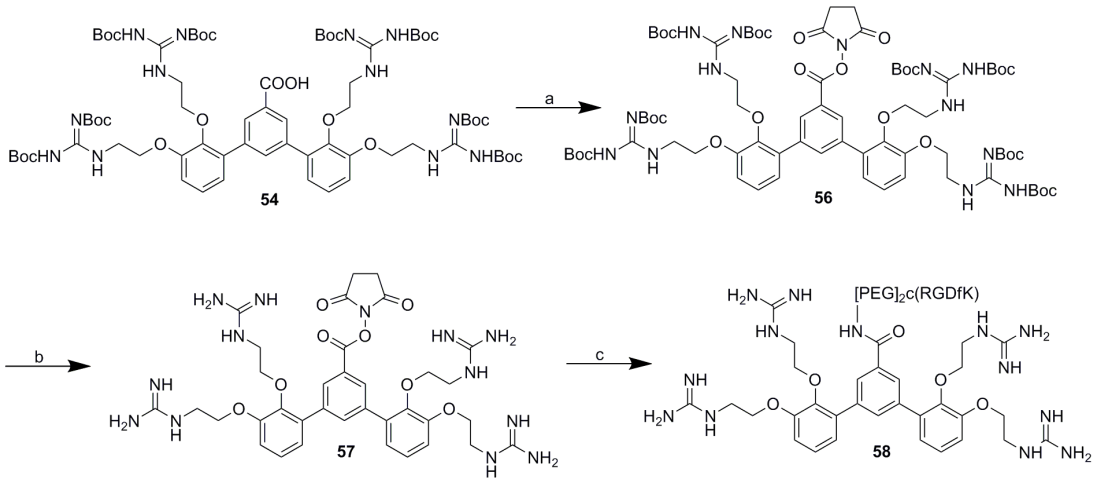
4.2.3. Conjugation to RGD

Initially, the coupling of 4G-SMoC-Dr to the peptide was attempted using the peptide coupling reagent HATU (Scheme 21), which activates the SMoC carboxylic acid for nucleophilic attack by the terminal amine of the peptide.[159] However, after stirring for 5 days with the PEG-linked RGD peptide in a solution of DMF and DIEA, followed by the addition of a further equivalent of HATU, no reaction was observed.



Scheme 21. Attempted direct coupling of cRGDfK to 4G-SMoC-Dr. Reagents: HATU, DIEA, DMF, rt, 5 days.

It was decided instead to synthesize and isolate an activated succinimidyl ester of the SMoC prior to coupling to the peptide (Scheme 22). The activated ester is also desirable, since it is stable in the absence of water, and therefore may be stored for future coupling reactions, which can be carried out in a single step. The activated ester was formed in 63% yield using *N*-hydroxysuccinimide (NHS) with the activating reagent 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), which forms an unstable intermediate which is attacked by the NHS nitrogen.[160] Since the succinimidyl ester is stable to acid, the guanidine groups were then deprotected using TFA, forming an activated form of 4G-SMoC-Dr in 50% yield which could be stored and easily coupled to peptides or other conjugates containing a terminal amino group. The coupling to cRGDfK was first attempted in an aqueous solution of $NaHCO_3$ at pH 8-9, however this resulted in hydrolysis of the ester without formation of product. This could be due to the low reactivity of the peptide terminal amine, which has low nucleophilicity due to the PEG chain which draws electrons away from the nitrogen. In aqueous conditions, the ester hydrolysis competes with nucleophilic attack, so the low nucleophilicity of the amine may result in the hydrolysis occurring preferentially. Therefore, the reaction was attempted in the polar aprotic solvent DMF using the organic base DIEA in order to reduce the solvation of the amine and increase its reactivity, as well as eliminating the competing ester hydrolysis reaction. This reaction was successful as observed by LC-MS, and the peptide-coupled product was obtained in 30% yield (Figure 34).



Scheme 22. Conjugation of 4G-SMoC-Dr to cyclic RGD peptide. Reagents: a) EDC, DMF, rt, 17 h (63%); b) TFA/TIPS/H_2O, rt, 3 h; c) cRGDfK, DIEA, DMF, rt, 17 h.

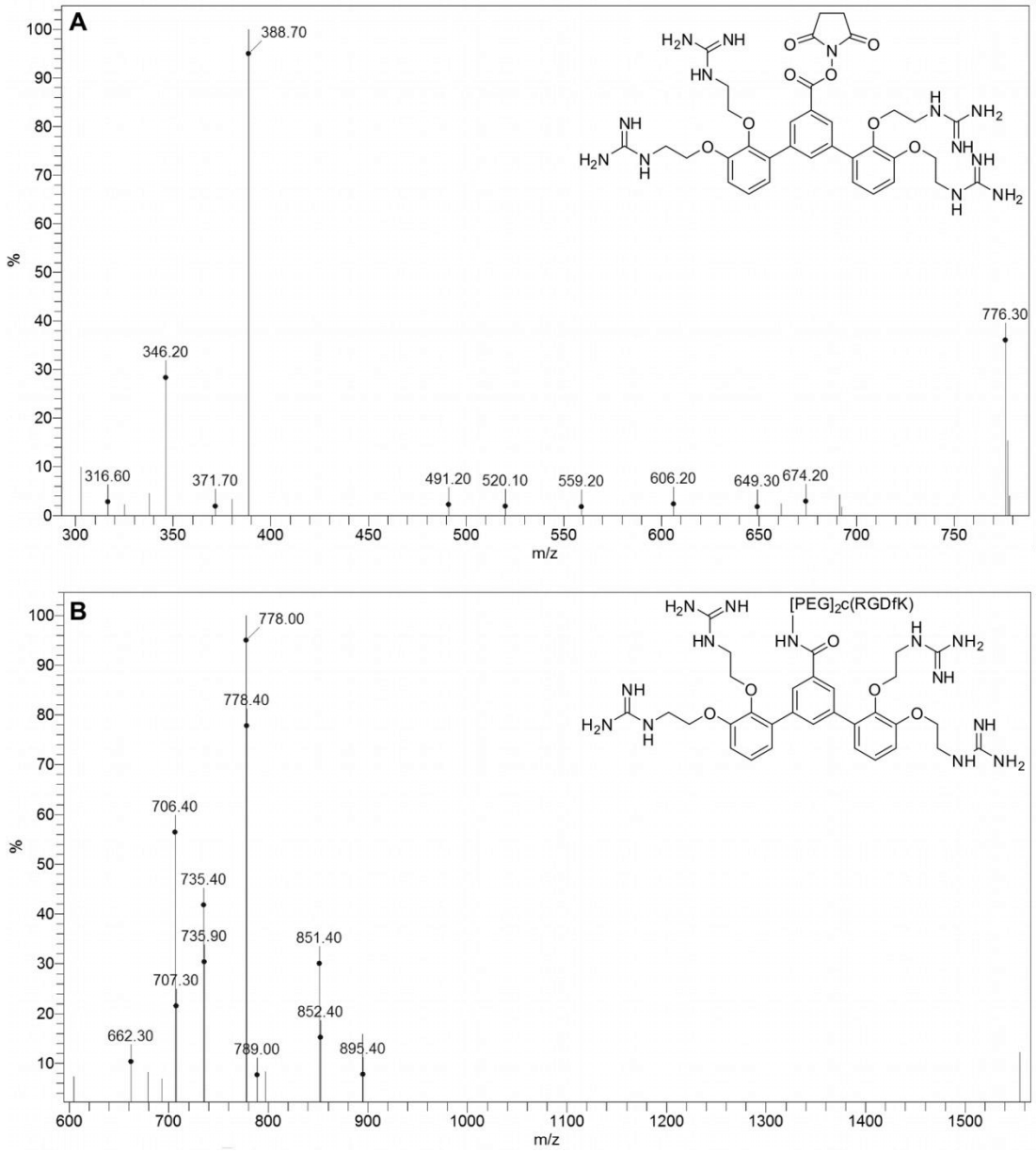


Figure 34. Monitoring of RGD coupling to 4G-SMoC-Dr by mass spectrometry. A: Activated ester **57** starting material. $[M + H]^+$ (*m/z* 776.3) and $[M + 2H]^{2+}$ (*m/z* 388.7) peaks observed; B: SMoC-RGD product **58**. $[M + H]^+$ (*m/z* 1554.7) and $[M + 2H]^{2+}$ (*m/z* 778.0) peaks observed.

4.2.4. Gel Shift

In order to confirm that the new SMoC retains the ability to bind siRNA, the binding affinity was assessed using a gel shift assay as described previously (Figure 35). The EC_{50} calculated was 0.07 mM, which is equivalent to 4G-SMoC-Py. This shows that

the dendrimer SMoC maintains the same level of siRNA-binding capabilities as its predecessor, which is expected since the new compound retains the same number of cationic guanidine groups, as well as the ability to form π-cation interactions with the central phenyl ring.

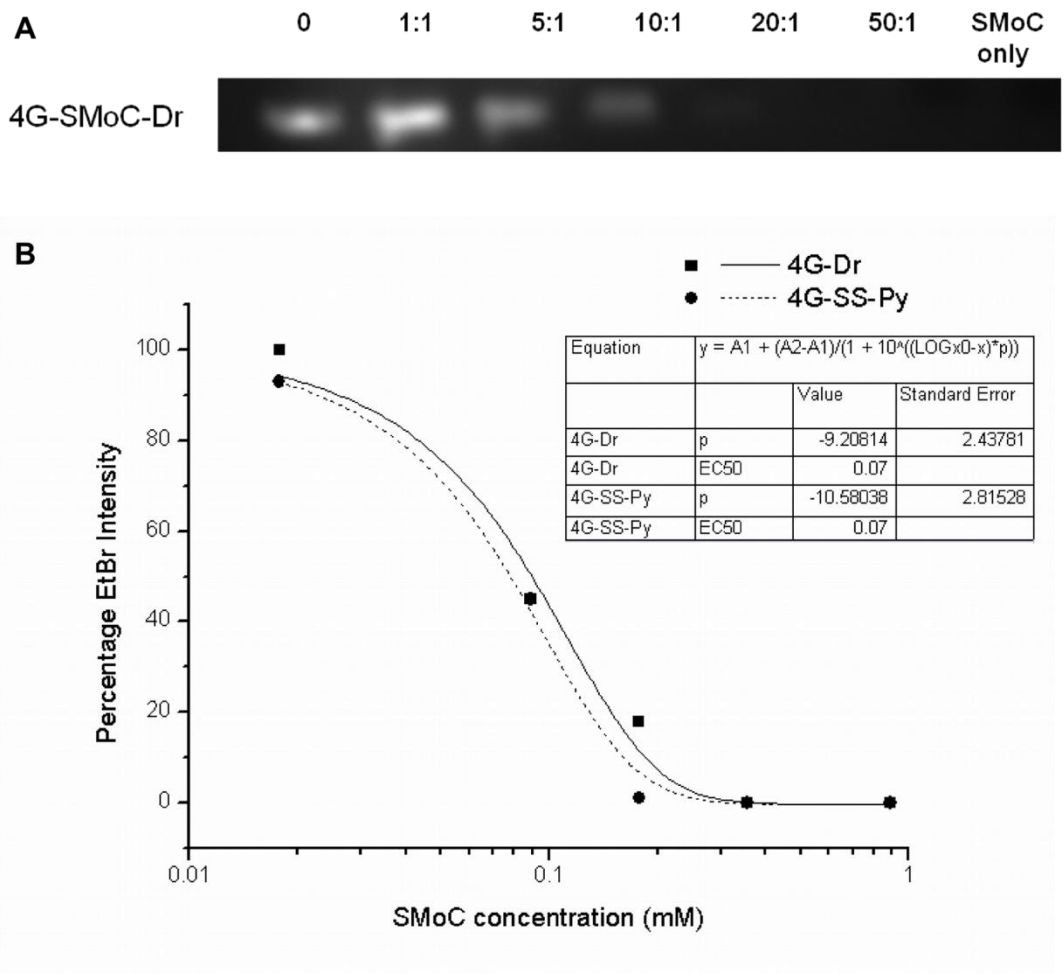


Figure 35. siRNA gel shift assay for 4G-SMoC dendrimer. A: GAPDH siRNA was mixed with 4G-SMoC-Dr at various ratios with a running dye containing EtBr and run on a 1% agarose gel. The gel was visualised under UV illumination. B: The EtBr intensities as a function of SMoC concentration with a binding curve fitted in Origin. The EC_{50} value and Hill coefficient (p) were calculated from the binding curve.

4.2.5. PCR

The new SMoC was used to transfect either *cdc7* siRNA or a negative control siRNA into IMR-90 primary human fibroblasts using the same protocol as described for the other SMoC compounds (Figure 36). Any knockdown of *cdc7* mRNA was measured using qRT-PCR and compared against the data obtained for the Lipofectamine-mediated transfection presented in the previous chapter. 4G-SMoC-Dr complexed with *cdc7* siRNA was successful in significantly reducing levels of *cdc7* mRNA in cells compared to the control siRNA, achieving a 69% knockdown, compared to 28% for Lipofectamine (Figure 37). However, the knockdown observed for the new compound was not statistically different from that of either Lipofectamine (P=0.7) or the original 4G-SMoC-Py (P=0.4) due to the large error value associated with the knockdown. Therefore, the new compound was not significantly more effective at delivering siRNA into cells than the compound described previously.

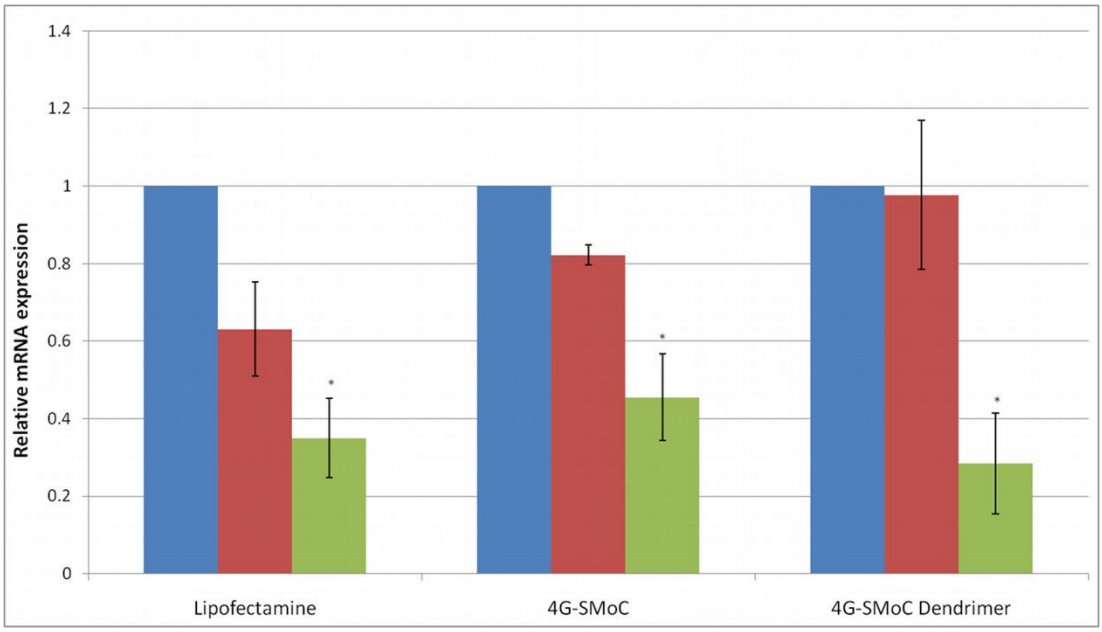


Figure 36. mRNA expression levels in cells treated with 4G-SMoC-Dr:siRNA complexes. qRT-PCR data showing *cdc7* mRNA expression in either untreated IMR-90 cells (blue) or cells treated with Lipofectamine, 4G-SMoC-Py, or 4G-SMoC dendrimer complexed to control (CO, red) or *cdc7* (green) siRNA. Experiments were conducted in triplicate. *, P<0.05 compared to CO siRNA.

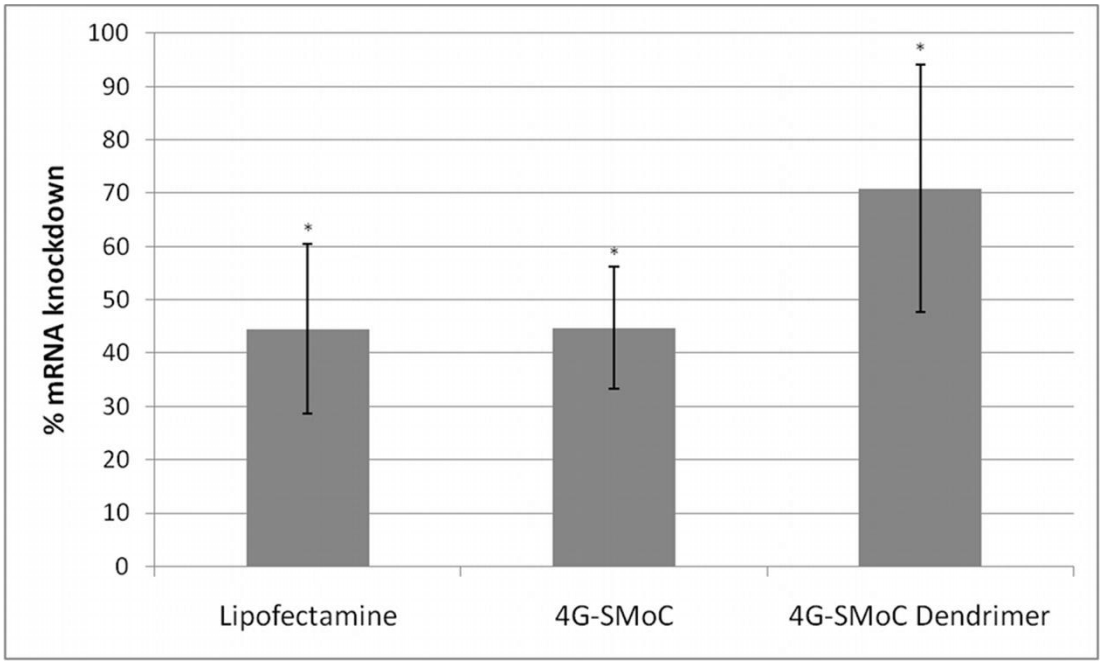


Figure 37. mRNA knockdown mediated by 4G-SMoC-Dr:siRNA complexes. c*dc7* mRNA knockdown in cells treated with *cdc7* siRNA + 4G-SMoC-Dr compared to the previous data collected for 4G-SMoC and Lipofectamine-mediated transfections. Knockdown was calculated as the difference in *cdc7* mRNA between cells transfected with the negative control siRNA and cells transfected with *cdc7* siRNA. Experiments were conducted in triplicate. *, $P<0.05$ compared to CO siRNA.

4.3. Conclusions

In the previous chapter, it was shown that the original 4G-SMoC-SSPy could act as an siRNA transfection agent, producing knockdowns of mRNA in IMR-90 and HEK-293 cells. Here, a new SMoC analogue has been designed in order to increase binding to cell surface proteoglycans, with the aim of increasing uptake efficiency by triggering internalisation via an endocytotic pathway. Using crystal structures of two GAGs the SMoC compound was designed to maximise binding to sulfate residues in order to achieve a high affinity to the cell surface. The compound was synthesized using the reactions optimised for previous SMoC compounds, with the additional observation that the guanidinylation reaction is possible in the presence of an acid linker. This new observation may be applied to future syntheses, for example in the production of the 4G-SMoC with a *meta* linker which was previously hindered by the

inability to hydrolyse the ester to form the acid linker. This synthesis has demonstrated that using the trifluoroborate salt developed earlier, a variety of new SMoC-related compounds may be easily synthesized via Suzuki couplings with aryl halides.

The interactions of the new compound, 4G-SMoC-Dr, with siRNA were characterised using the gel shift assay used for previous compounds, and a binding affinity calculated in order to compare it to the other SMoCs. As hypothesized, the new compound retained the ability to complex siRNA displaying a similar binding affinity to the original 4G-SMoC-SSPy, showing that the addition of a phenyl ring as well as the presence of an acid does not affect its ability to complex siRNA. Finally, IMR-90 cells were successfully transfected with 4G-SMoC-Dr, producing a significant RNAi-induced knockdown of the target mRNA. The knockdown achieved was not significantly different from the previous 4G-SMoC compound, or from Lipofectamine. Therefore, the hypothesis that additional branching of the SMoC side chains will increase internalisation via GAG clustering is not supported by this data.

Using an activated ester of the 4G-SMoC-Dr compound, it was possible to attach the RGD targeting peptide as predicted. The activated ester of 4G-SMoC-Dr is a significant development in SMoC chemistry as it allows the simple attachment of targeting ligands, such as peptides or antibodies, in order to target specific cell types or cell receptors which may improve uptake. In addition, as seen in the previous chapter, the inclusion of a disulfide-containing linker group may increase uptake due to either dimerisation or by interaction with cell surface thiols. Therefore inclusion of this type of functionality may improve the efficiency of siRNA internalisation for 4G-SMoC-Dr.

CHAPTER 5

Experimental Methods

5. Experimental Methods

5.1. General Methods

5.1.1. General chemistry methods

Starting materials were either commercially available or synthesized according to methods reported in the literature. ^{1}H and ^{13}C NMR spectra were recorded on either a Bruker AMX-300 or a Bruker AMX-500 spectrometer. Chemical shifts are reported as ppm relative to TMS internal standard. Mass spectra were recorded on either a Fisons VG70-SE spectrometer (EI, FAB) or an Agilent 6100 Series LC-mass spectrometer using either a C-18 or C-4 column. Microwave reactions were carried out using a CEM Discover microwave. All compounds were purified to ≥95% as measured by LC-MS.

5.1.2. Gel shifts

Polyarginine peptides were obtained from Peptide Protein Research, Hampshire, UK. 1.7 nmol GAPDH siRNA (Dharmacon RNAi Technologies; sequences used were sense 5' UGG UUU ACA UGU UCC AAU AUU 3', antisense 5' Phosphate U AUU GGA ACA UGU AAA ACC UU 3') was mixed in RNAse-free water with varying concentrations of either SMoC, tetraarginine or octaarginine in the molar ratios of 1/1, 1/5, 1/10, 1/20 and 1/50. The complexes were incubated for 30 min at room temperature. The samples were diluted with 2x RNA Loading Dye (Fermentas UK) containing ethidium bromide and analysed by gel electrophoresis on a 1% agarose gel. The gels were analysed by UV illumination and the ethidium bromide bands quantified using the ImageJ software. The EtBr intensities were plotted, a dose response curve fitted and EC_{50} values calculated using the Origin software.

5.1.3. Cell Culture

IMR-90 (ATCC# CCL-186), a diploid primary human fibroblast adherent cell line derived from fetal lung tissue, was obtained from LGC Standards (Middlesex, UK) at population doubling (PD) 12. HEK-293 (ATCC# CRL-1573) cells were kindly donated by Neale Foxwell (Wolfson Institute of Biomedical Research, UK). Cells

were cultured in DMEM (Invitrogen, UK) supplemented with 10% Fetal Bovine Serum (Invitrogen, UK) and 2 mM L-glutamine (Invitrogen, UK) at 37 °C and 5% CO_2.

5.1.4. siRNA complexes

Solutions were made up in Opti-MEM® (Invitrogen, UK); 100 nM *cdc7* siRNA (sense 5'-GCU CAG CAG GAA AGG UGU UUU-3', antisense 5'-AAC ACC UUU CCU GCU GAG CUU-3'; Applied Biosystems, UK), 100 nM Accell Non-targeting siRNA 1 negative control siRNA (Dharmacon RNAi Technologies), 8 µl/ml Lipofectamine LTX reagent (Invitrogen, UK), 0.5 mM SMoC. SMoC/siRNA and Lipofectamine/siRNA complexes were made by mixing equal volumes of the required solutions and incubating for 30 min at room temperature.

5.1.5. siRNA Transfection

Cells were plated in 6-well plates at a density of 5 x 10^5 cells/well (IMR-90 cells) or 1 x 10^5 cells/well (HEK-293 cells) in 1 ml DMEM containing 10% FCS and 2 mM glutamine and incubated at 37 °C under 5% CO_2 for 24 h. The media was removed and replaced with 1.6 mL Opti-MEM® (Invitrogen, UK). 400 µl of SMoC/siRNA or Lipofectamine/siRNA complexes were added and the cells incubated for a further 72 h. The cells were washed with PBS and trypsinised and the pellet stored at -80°C.

5.1.6. RNA extraction and qRT-PCR

RNA was extracted from the cell pellets using the PureLink RNA Mini Kit (Invitrogen, UK) using the protocol provided by the manufacturer. RNA was quantified using a Nanodrop spectrophotometer and 40 ng of RNA was mixed with Superscript III RT / Platinum *Taq* Mix and SYBR Green Reaction Mix from the Supersript III Platinum SYBR Green One-Step qRT-PCR Kit (Invitrogen, UK) according with the manufacturer's protocol. PCR conditions were 50 °C for 3 min, 95 °C for 10 min (1 cycle), 95 °C for 15 s, 47 °C for 30 s, 60 °C for 30 s (45 cycles), 95 °C for 15 s, 60 °C for 15 s, ramp to 95 °C over 20 min (1 cycle). Relative quantitation data were obtained using the comparative C_t method with Realplex

software according to the manufacturer's protocol (Eppendorf, Heidelberg, Germany). c*dc7* RNA was measured using the following primers: forward primer: 5'-AAC TTG CAG GTG TTA AAA AAG-3': reverse primer: 5'-TGA AAG TGC CTT CTC CAA T-3'. The following primers were used to measure GAPDH: forward primer: 5'-TCA ACT ACA TGG TTT ACA TGT TC-3': reverse primer: 5'-GAT CTC GCT CCT GGA AGA T-3'. The signal obtained from the c*dc7* primers was normalised using the signal from the GAPDH primers.

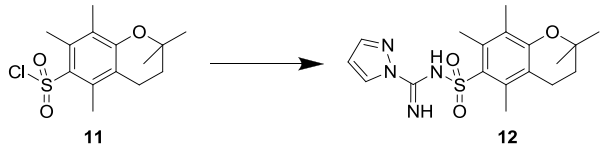
5.1.7. Statistical Analysis

Groups were compared using an unpaired two-sample Student's t-test. A value of $P<0.05$ was considered significant.

5.2. Chapter 2 Methods

5.2.1. Syntheses

5.2.1.1. *N-(2,2,5,7,8-pentamethylchroman-6-ylsulfonyl)-1H-pyrazole-1-carboximidamide*[131] **12**



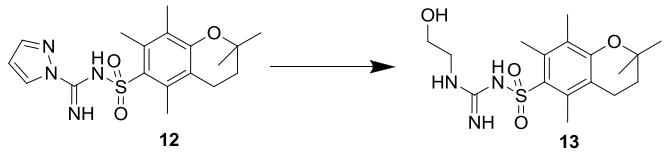
To a solution of 2,2,5,7,8-pentamethylchroman-6-sulfonyl chloride **11** (1.06 g, 3.5 mmol) in dioxane (30 mL) was added 1*H*-pyrazole-1-carboxamidine (0.52 g, 3.5 mmol) and DIEA (1.15 mL, 7 mmol) and the mixture stirred at 50 °C for 17 h. The dioxane was removed under vacuum and ethyl acetate and water were added in a 1:1 ratio until the mixture was completely dissolved. The layers were separated and the organic layer washed with water (2 x 20 mL), dried over $MgSO_4$, filtered, and the solvent removed under vacuum to afford **12** (0.56 g, 43% yield) as a pale yellow solid. The compound was not purified further for the next stage.

^{1}H-NMR (300 MHz, $CDCl_3$) δ: 1.31 (s, 6H, OC(*C*H_3)$_2$), 1.82 (t, *J* = 6.8 Hz, 2H, O-C(CH$_3$)$_2$*C*H_2), 2.12 (s, 3H, Ar-*C*H_3), 2.59 (m, 8H, Ar-*C*H_3 + *C*H_2), 6.40 (dd, *J* = 1.62, 2.73 Hz, 1H, c(-NCH*CH*CHN-)), 7.40 (br, 1H, *NH*), 7.66 (m, 1H, N-N-*CH*), 7.84 (br, 1H, NH), 8.21 (d, *J* = 2.8 Hz, 1H, C-N-*CH*).

^{13}C-NMR (300 MHz, $CDCl_3$) δ: 12.1 (CH_3), 18.5 (CH_3), 21.4 (CH_3), 26.8 (2CH_3), 26.9 (CH_2), 32.7 (CH_2), 74.0 (*C*(CH$_3$)$_2$, 109.6 (-NCH*CH*CHN-), 118.2 (Ar*C*), 124.5 (Ar*C*), 129.1 (C-N-*CH*), 135.8 (Ar*C*), 143.6 (N-N-*CH*), 148.8 (Ar*C*), 154.5 (*C*=N).

LRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{18}H_{25}N_4O_3S$, 377.2; found, 377.4.

5.2.1.1. *N-(N-(2-hydroxyethyl)carbamimidoyl)-2,2,5,7,8-pentamethylchroman-6-sulfonamide* **13**

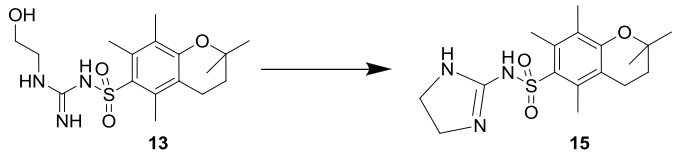


Ethanolamine (1.05 g, 17.5 mmol) was dissolved in dioxane (17.5 mL) and a solution of **12** (0.5 g, 1.49 mmol) in dioxane (10 mL) was added dropwise to the stirred solution and the reaction mixture stirred overnight. The solvent was removed under vacuum and ethyl acetate and water were added. The layers were separated and the organic layer was washed with water (2 x 20 mL), dried over $MgSO_4$, filtered, and concentrated under vacuum. The residue was purified by flash column chromatography using DCM/methanol (9/1) as eluent to afford **13** as a white solid (0.22 g, 40 % yield).

^{1}H-NMR (300 MHz, $(CD_3)_2CO$) δ: 1.29 (s, 6H, OC(*CH*$_3$)$_2$), 1.82 (t, *J* = 6.9 Hz, 2H, O-C(CH$_3$)$_2$*CH*$_2$), 2.03 (m, 3H, Ar-*CH*$_3$), 2.55 (d, *J* = 5.3 Hz, 6H, 2 x Ar-*CH*$_3$), 2.64 (t, J = 6.8 Hz, 2H, O-C(CH$_3$)$_2$CH$_2$*CH*$_2$), 3.28 (m, 2H, NH*CH*$_2$), 3.58 (m, 2H, OH*CH*$_2$).

LRMS-ES (*m/z*): $[M + H]^+$ calcd. for $C_{17}H_{28}N_3O_4S$, 370.2; found, 370.4.

5.2.1.2. *N-(4,5-dihydro-1H-imidazol-2-yl)-2,2,5,7,8-pentamethylchroman-6-sulfonamide* **15**



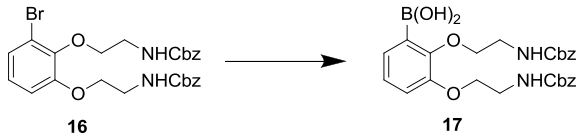
N-(N-(2-hydroxyethyl)carbamimidoyl)-2,2,5,7,8-pentamethylchroman-6-sulfonamide **13** (100 mg, 0.27 mmol) was dissolved in DCM (10 mL) and methanesulfonyl chloride (1 or 2 eq) was added followed by triethylamine (0.037 ml, 0.27 mmol) or anhydrous sodium carbonate (28 mg, 0.27 mmol). The mixture was stirred for 5 h and monitored by LC-MS.

LRMS-ES (*m/z*): $[M + H]^+$ calcd. for $C_{17}H_{26}N_3O_3S$, 351.5; found, 351.4.

5.2.1.3. *Mitsunobu General Method*

The required phenol (100 mg) was stirred at 0°C with 1 equivalent of phosphine in THF (25 ml) for 5 mins. Alcohol **13** (1 eq) and the Mitsunobu coupling reagent (TMAD or DIAD) (1 or 2 eq) were added and the reaction warmed to room temperature and stirred for 24 h. The reaction was monitored by LC/MS.

5.2.1.4. *2,3-Bis(2-benzyloxycarbonylamino)ethoxy)phenylboronic acid* **17**



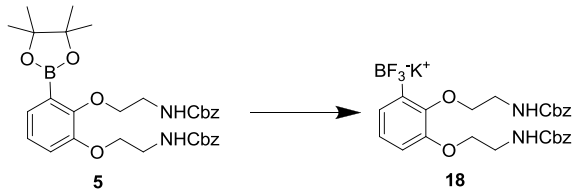
16 **17**

To a solution of bis(2-(*N*,*N*-dimethylamino)ethyl) ether (3.67 mL, 20 mmol) in anhydrous THF (25 mL) under N_2 at 10-15 °C was added ^{i}PrMgCl (2 M solution in THF, 10 mL, 20 mmol). The mixture was stirred at this temperature for 15 min then allowed to warm to room temperature. Dibenzyl 2,2'-(3-bromo-1,2-phenylene)bis-(oxy)bis(ethane-2,1-diyl)dicarbamate **16** (5.43g, 10 mmol) in anhydrous THF (25 mL) was added and the reaction mixture stirred at room temperature under N_2. Consumption of starting material was monitored by LC-MS. After 4 h ^{i}PrMgCl (2 M solution in THF, 5 mL, 10 mmol) was added and the reaction mixture stirred at room temperature under N_2 overnight. Trimethyl borate (4.60 mL, 40 mmol) was added and the reaction mixture stirred for 1 h. 0.1 M HCl (aq) (150 mL) was added and the product extracted with EtOAc (3 x 250 mL). The combined organic fractions were dried over $MgSO_4$, filtered and concentrated under vacuum. The product was purified by flash chromatography using a gradient of 0-100% EtOAc in cyclohexane over 12 CV to afford the title product (2.05 g, 40%) as a colourless oil.

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.27-3.71 (m, 4H, $2CH_2$), 3.84–4.18 (m, 4H, $2CH_2$), 4.94-5.22 (m, 4H, 2Cbz-CH_2), 5.48 (s, 1H, NH), 5.90 (s, 1H, NH), 6.79-7.18 (m, 2H, 2Ar-CH), 7.19–7.61 (m, 11H, 11Ar-CH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 40.7 (CH_2), 41.3 (CH_2), 67.0 (CH_2), 67.3 (CH_2), 67.8 (CH_2), 116.5 (Ar-CH), 125.2 (Ar-CH), 128.0 (Cbz Ar-CH), 128.3 (Cbz Ar-CH), 128.6 (Cbz Ar-CH), 136.3 (2 x Cbz Ar-C), 150.5 (Ar-*C*-OCH_2CH_2), 152.4 (Ar-*C*-OCH_2CH_2), 156.6 (2 x C=O).

5.2.1.5. *Potassium 2,3-Bis(2-benzyloxycarbonylamino)ethoxy)phenyltrifluoroborate* **18**



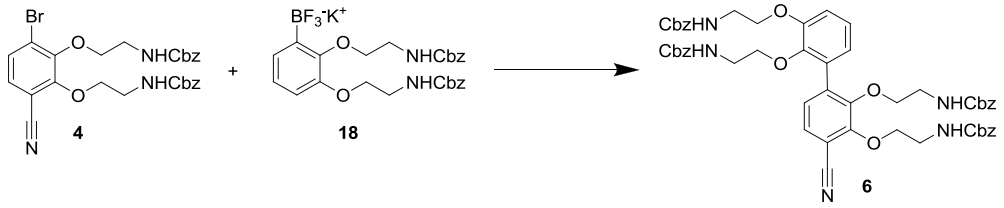
To a stirred solution of benzyl 2,2'-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1,2-phenylene)bis(oxy)bis(ethane-2,1-diyl)dicarbamate (pinacol boronate) **5** (5.66 g, 9.6 mmol) in MeOH (50 ml) was added dropwise a solution of KHF_2 (6.0 g, 76.72 mmol) in H_2O (17.0 ml). The clear yellow solution formed a white precipitate after 5 minutes. The resulting mixture was stirred at room temperature for 3 h. The mixture was then filtered and the white solid residue washed with DCM (3 x 20 ml), cold acetone (2 x 20 ml) and hot acetone (2 x 20 ml). The washings and the filtrate were combined and the solvent removed under vacuum to leave a brown solid which was further purified by washing several times with diethyl ether to leave the salt (4.85 g, 8.40 mmol, 88%) as a white solid.

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.07- 3.35 (m, 4H, 2 x C*H*$_2$NH), 3.77 (bs, 4H, 2 x C*H*$_2$O), 5.19 (s, 4H, 2 x ArC*H*$_2$O), 5.70 (bs, 1H, N*H*), 5.77 (bs, 1H, N*H*), 6.57 (d, *J* = 6.57 Hz, 1H, Ar*H*), 6.73 (t, *J* = 7.2 Hz, 1H, Ar*H*), 6.97 (d, *J* = 5.5, 1H, Ar*H*), 7.34 (m, 11H).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 40.7 (NH-*C*H_2), 40.8 (NH-*C*H_2), 66.7 (Ph-*C*H_2), 67.6 (2 x O-*C*H_2CH_2), 124.4 (Ar*C*-B), 127.9 (Ar*C*), 128.0 (Ar*C*), 128.1 (Ar*C*), 128.4 (Ar*C*), 128.5 (Ar*C*), 136.6 (Ar*C*-CH_2), 150.7 (Ar*C*-O), 156.8 (Ar*C*-O), 157.3 (*C*=O).

Anal. Calcd. for $C_{26}H_{27}BF_3KN_2O_6$: C 54.75 %, H 4.77%, N 4.91%, B 1.90%, F 9.99%, K 6.85 %. Found: C 54.19%, H 4.80%, N 4.77%, B 1.92%, F 10.22%, K 6.36%.

5.2.1.6. *Tetrabenzyl 2,2',2'',2'''-(4-cyanobiphenyl-2,2',3,3'-tetrayltetrakis(oxy))tetrakis-(ethane-2,1-diyl)tetracarbamate* **6**

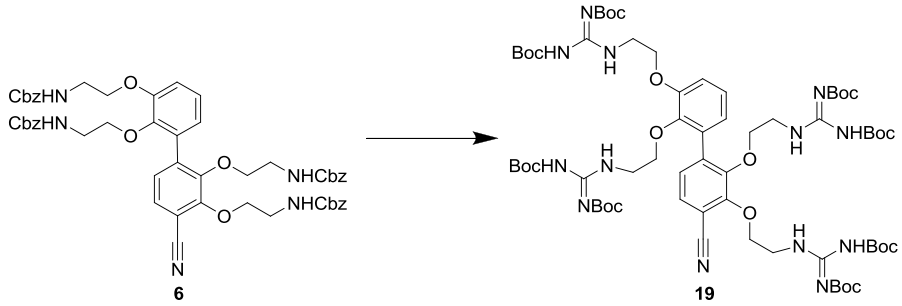


To a degassed solution of potassium (2,3-bis(2-(benzyloxycarbonylamino)ethoxy)-phenyl) trifluoroborate **18** (800 mg, 1.40 mmol) in isopropanol/water (2:1) (26 mL) was added $PdCl_2dppf.CH_2Cl_2$ (40 mg, 0.049 mmol) and triethylamine (600 µL, 4.2 mmol). The mixture was stirred at room temperature for 2 minutes then dibenzyl 2,2'-(3-bromo-6-cyano-1,2-phenylene) bis(oxy)bis(ethane-2,1-diyl)dicarbamate **4** (400 mg, 0.70 mmol) was added. The reaction mixture was stirred at 82 °C for 18 h. The reaction mixture was diluted with 1M HCl (aq) (50 mL) and the product extracted with DCM (3 x 200 mL). The combined organic fractions were dried over $MgSO_4$, filtered and concentrated under reduced pressure. The product was purified by flash chromatography using a gradient of 0-50% EtOAc in cyclohexane over 10 CV to afford the title product (530 mg, 79%) as a white foam.

^{1}H-NMR (300 MHz, $CDCl_3$) δ: 3.04-3.34 (m, 4H, 2 x CH_2), 3.51 (s, 4H, 2 x CH_2), 3.69-3.36 (m, 8H, 4 x CH_2), 4.89-5.20 (m, 8H, 4 x COOCH_2Ar), 6.68-6.92 (m, 2H, 2 x ArH), 6.92-7.14 (m, 2H, 2 x ArH), 7.12-7.44 (m, 20H, 20 x Cbz-ArH).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{53}H_{54}N_5O_{12}$, 952.37688; found, 952.37462.

5.2.1.7. *2, 3, 2', 3'-Tetra(2-(N, N'-bis(tert-butoxycarbonyl)guanidino)-ethyloxy)-4-cyanobiphenyl* **19**

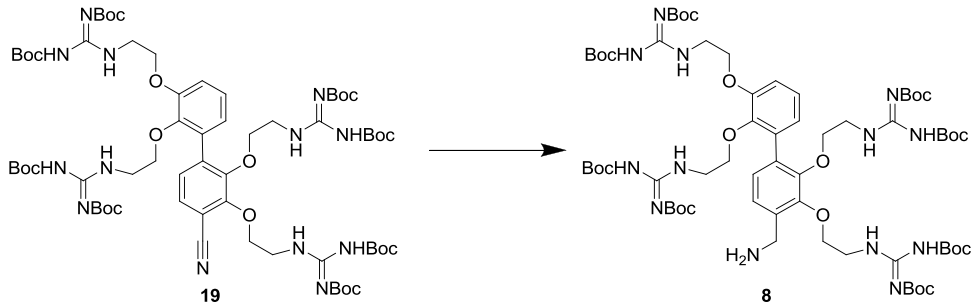


To a solution of benzyl 2,2',2'',2'''-(4-cyanobiphenyl-2,2',3,3'-tetrayl)tetrakis(oxy)-tetrakis(ethane-2,1-diyl)tetracarbamate **6** (550 mg, 0.58 mmol) in DCM (25 mL) was added a 30% solution of HBr in acetic acid (5.5 mL) dropwise. The reaction mixture was stirred at room temperature for 1 h then water (20 mL) was added and the layers separated. The organic fraction was extracted with water (20 mL). The combined aqueous fractions were concentrated under reduced pressure and the crude product was dried under vacuum for 8 hours. The resulting solid was suspended in DCM (25 mL) and DIEA (1 mL, 5.84 mmol) was added. After stirring for 10 min *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine (790 mg, 2.54 mmol) was added and the reaction mixture stirred at room temperature for 18 h. The reaction mixture was diluted with DCM (50 mL) then washed with 10% citric acid (aq) (25 mL), followed by sat. aq. $NaHCO_3$ (25 mL) and brine (25 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated under vacuum. The product was purified by flash chromatography using a gradient of 0-50% EtOAc in cyclohexane over 12 CV to afford the title product (570 mg, 71%) as a white foam.

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.34-1.58 (m, 72H, 24 x Boc-CH_3), 3.43-3.64 (m, 4H, 2 x CH_2), 3.79-4.01 (m, 8H, 4 x CH_2), 4.09-4.23 (m, 2H, CH_2), 4.33-4.48 (m, 2H, CH_2), 6.88-7.08 (m, 3H, 3 x Ar-H), 7.19 (d, *J* = 8.1 Hz, 1H, Ar-H), 7.30 (m, 1H, Ar-H).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{65}H_{102}N_{13}O_{20}$, 1384.73637; found, 1384.73912.

5.2.1.8. *2, 3, 2', 3'-Tetra(2-(N, N'-bis(tert-butoxycarbonyl)guanidino)-ethyloxy)-4-(aminomethyl)biphenyl* **8**

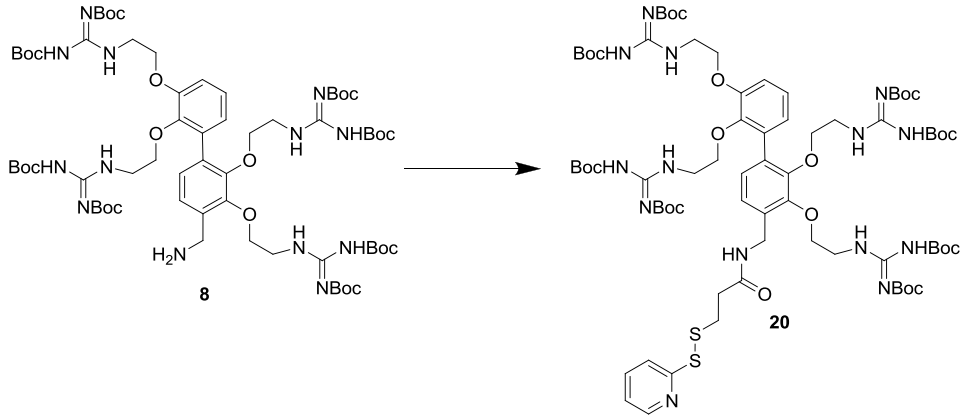


To a solution of 2,3,2',3'-tetra(2-(*N*, *N*'-bis(tert-butoxycarbonyl)guanidino)-ethyloxy)-4-cyanobiphenyl **19** (400 mg, 0.290 mmol) in MeOH (8 mL) was added $CoCl_2.6H_2O$ (136 mg, 0.576 mmol) followed by $NaBH_4$ (112 mg, 2.90 mmol). The reaction mixture was stirred at room temperature for 1 h. Water (25 mL) and EtOAc (50 mL) were added and the layers separated. The aqueous layer was extracted with EtOAc (2 x 50 mL). The combined organic fractions were dried over $MgSO_4$, filtered and evaporated. The product was purified by flash chromatography using a gradient of 50-100% EtOAc in cyclohexane to afford the title product (240 mg, 60%) as a white foam.

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.17-1.66 (m, 72H, 24 x Boc-CH_3), 3.31-3.61 (m, 4H, 2 x CH_2), 3.68-4.02 (m, 8H, 4 x CH_2), 4.04-4.33 (m, 4H, 2 x CH_2), 4.36-4.50 (m, 2H, CH_2NH_2), 6.80-7.21 (m, 5H, 5 x ArH).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{65}H_{106}N_{13}O_{20}$, 1388.7677; found, 1388.7656.

5.2.1.9. *tert-Butyl (2,2',2'',2'''-(4-((3-(pyridine-2-yldsulfanyl)propanamido)methyl)-biphenyl-2,2',3,3'-tetrayl)tetrakis(oxy)tetrakis(ethane-2,1-diyl)tetrakis(azanediyl)-tetrakis((tert-butoxycarbonylamino)methane-1-yl-1-ylidene)tetracarbamate* **20**

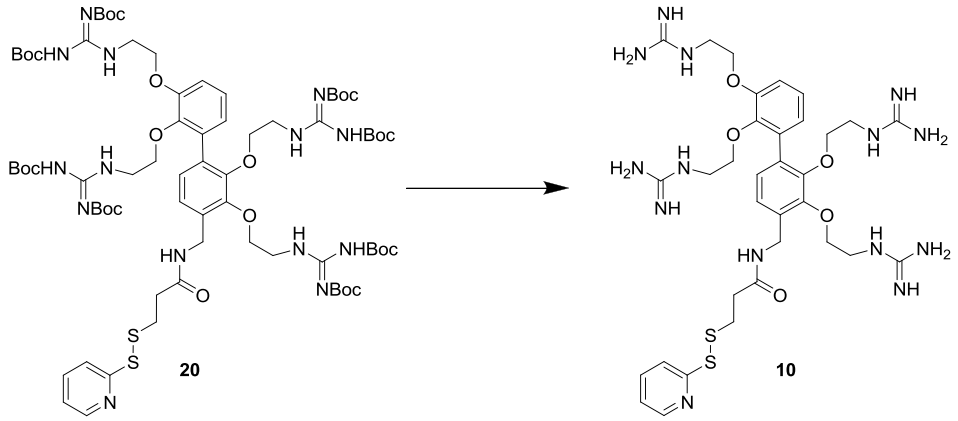


To a solution of 2,3,2',3'-tetra(2-(*N*, *N'*-bis(tert-butoxycarbonyl)guanidino)-ethyloxy)-4-(aminomethyl)biphenyl **8** (140 mg, 0.101 mmol) and DIEA (42 µL, 0.248 mmol) in DCM (1.5 mL) was added 2,5-dioxopyrrolidine-1-yl-3-(pyridine-2-yldisulfanyl)propanoate (39 mg, 0.124 mmol). The reaction mixture was stirred at room temperature in the dark for 18 h. Solvent was removed under reduced pressure and the product was purified by flash chromatography using a gradient of 0-100% EtOAc in cyclohexane to afford the title product (75 mg, 57%) as a pale yellow oil.

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.34-1.60 (m, 72H, 24 x Boc-CH_3), 2.71-2.92 (m, 4H, 2 x CH_2), 3.50-3.78 (m, 5H), 3.80-3.94 (m, 5H), 4.02-4.52 (m, 10H), 6.86-7.21 (m, 6H, 6 x ArH), 7.28-7.81 (m, 3H, 3 x ArH).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{73}H_{113}N_{14}O_{21}S_2$, 1585.7646; found, 1585.7542.

5.2.1.10. *2,2',2'',2'''-(4-((3-(Pyridine-2-yldsulfanyl)propanamido)methyl)biphenyl-2,2',3,3'-tetrayl)tetrakis(oxy)tetrakis(ethane-2,1-diyl)tetrakis(azanediyl)tetrakis-(aminomethaniminium)* **10**

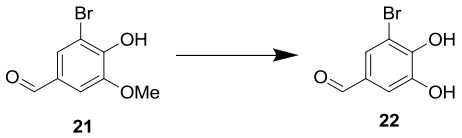


To a solution of **20** (40 mg, 0.0252 mmol) in DCM (2 mL) was added a solution of TFA/*m*-cresol (95:5) (2 mL). The reaction mixture was stirred at room temperature for 3 h. Solvent was evaporated and the product precipitated from Et_2O, washed three times with Et_2O, dissolved in water and lyophilised to give the title product (28mg, 82%) as a white powder.

^{1}H-NMR (500 MHz, MeOD) δ: 2.73 (t, *J* = 6.8 Hz, 2H, *CH*$_2$CH$_2$SS), 3.10 (t, *J* = 6.4 Hz, 2H, *CH*$_2$SS), 3.51-3.79 (m, 4H, 2 x CH_2), 3.78-4.08 (m, 8H, 4 x CH_2), 4.14-4.37 (m, 4H, 2 x CH_2), 6.75-7.32 (m, 9H, 9 x ArH).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{33}H_{49}N_{14}O_5S_2$, 785.3452; found, 785.3442.

5.2.1.11. *3-bromo-4,5-dihydroxybenzaldehyde* **22**[138]



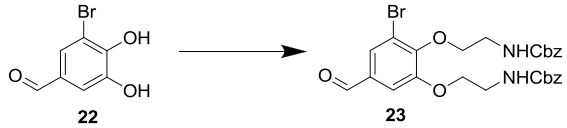
5-bromovanillin **21** (15 g, 65 mmol) was dissolved in chloroform (150 mL) and stirred at 0°C for 15 min. $AlCl_3$ (12.4 g, 93 mmol) was added and pyridine (23 mL) was added dropwise while stirring at 0°C. A reflux condenser was fitted and the green solution heated to reflux and stirred for 24 h. The dark mixture was allowed to cool and the solvent evaporated under vacuum. 3M HCl was added until the mixture was acidic, and the solid that precipitated was washed with water and filtered. The dark brown product was purified by flash column chromatography (1:1 acetone:cyclohexane) to yield a brown solid (8.36 g, 80% yield).

^{1}H-NMR (500 MHz, $(CD_3)_2CO$) δ: 7.39 (d, *J* = 1.8 Hz, 1H, Ar*H*), 7.64 (d, *J* = 1.8 Hz, 1H, Ar*H*), 9.24 (br, 2H, 2 x O*H*), 9.79 (s, 1H, *HC*=O).

^{13}C-NMR (500 MHz, $(CD_3)_2CO$) δ: 109.5 (Ar*C*-Br), 113.5 (Ar*C*H), 127.9 (Ar*C*H), 130.6 (Ar*C*-CHO), 146.4 (Ar*C*-OH), 149.4 (Ar*C*-OH), 189.9 (*C*=O).

HRMS-EI (*m/z*): $[M - H]^-$ calcd for $C_7H_4{}^{79}BrO_3$, 214.93383; found, 214.93377.

5.2.1.12. *Benzyl 2,2'-(3-bromo-5-formyl-1,2-phenylene)bis(oxy)bis(ethane-2,1-diyl)dicarbamate* **23**



3-bromo-4,5-dihydroxybenzaldehyde **22** (1.46 g, 6.74 mmol) was dissolved in DMF (30 mL) with Na_2CO_3 (3.57 g, 33.7 mmol) and stirred under a nitrogen atmosphere at 80°C for 10 min. Mesylate **3** (7.37 g, 27 mmol) was added and the reaction stirred for 17 h. The solvent was evaporated under vacuum and the residue dissolved in ethyl acetate (50 mL) and washed with water (3 x 50 mL). The organic layer was dried over $MgSO_4$, filtered and the solvent removed under vacuum. The mixture was purified by flash chromatography and the product isolated as a yellow oil (2.42 g, 46% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.50–3.59 (m, 4H, 2 x *CH*$_2$NH), 4.14 (t, *J* = 4.7 Hz, 4H, 2 x O*CH*$_2$CH$_2$), 5.07 (s, 4H, 2 x Ph-*CH*$_2$), 5.70 (br, 2H, 2 x NH), 7.25-7.61 (m, 12H, Ar-H), 9.78 (s, 1H, CHO).

^{13}C-NMR (500 MHz, $CDCl_3$) $_{\delta}$: 40.4 (CH_2) 40.6 (CH_2), 66.6 (CH_2), 68.1 (CH_2), 71.9 (CH_2), 109.9 (Ar-CH), 126.9 (Ar-CH), 128.0 (Cbz-ArCH), 128.2 (Cbz-ArCH), 130.4 (Ar-C), 136.6 (Ar-C), 146.9 (Ar-C), 152.4 (Ar-C), 156.6 (Ar-C), 175.0 (Ar-C=O).

HRMS-ES (*m/z*): $[M + Na]^+$ calcd for $C_{27}H_{27}N_2O_7NaBr$, 593.0899; found, 593.0891.

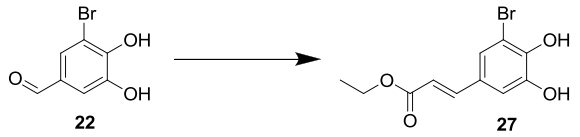
5.2.1.13. *Nitrile synthesis general method*

Benzyl 2,2'-(3-bromo-5-formyl-1,2-phenylene)bis(oxy)bis(ethane-2,1-diyl)dicarbamate **23** (100 mg, 0.18 mmol) or 3-bromo-4,5-dihydroxybenzaldehyde **22** (100 mg, 0.46 mmol) was dissolved in NMP or DMF (5 mL) and hydroxylammonium chloride (1.5 eq) was added. The mixture was either heated by microwave irradiation or thermally (Table 6) and the reaction monitored by LC-MS. The reaction mixture was concentrated under vacuum, dissolved cyclohexane/ethyl acetate (1:1) and

purified by flash column chromatography using a cyclohexane/ethyl acetate gradient (0-50%). The oxime derived from **23** was isolated and characterised by mass spectrometry.

HRMS-ES (*m/z*): $[M + Na]^+$ calcd for $C_{27}H_{28}BrN_3O_7Na$, 610.1090; found, 610.1059.

5.2.1.14. *(E)-ethyl 3-(3-bromo-4,5-dihydroxyphenyl)acrylate* **27**



Method 1: 3-bromo-4,5-dihydroxybenzaldehyde **22** (100 mg, 0.46 mmol) and ethyl malonate (0.14 mL, 1.15 mmol) were dissolved in pyridine (5 mL) and piperidine (0.16 mL) and the mixture stirred at room temperature for 17 h. The reaction progress was monitored by LC-MS.

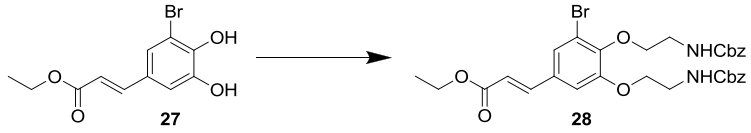
Method 2: To a stirring solution of 3-bromo-4,5-dihydroxybenzaldehyde **22** (217 mg, 1 mmol) in DCM (3 mL) was added (carbethoxymethylene)triphenylphosphorane (418 mg, 1.2 mmol) and the mixture stirred at room temperature for 17h. The solvent was evaporated and the product mixture absorbed onto silica and purified by flash column chromatography using a 0-50% ethyl acetate gradient in cyclohexane. The final product was recrystallised from a mixture of diethyl ether and cyclohexane (1:1) to yield the title compound as a dark yellow solid (175.7 mg, 76%).

^{1}H-NMR (500 MHz, $(CD_3)_2CO$) δ: 1.29 (t, *J* = 7.0 Hz, 3H, CH_3), 2.88 (br, 2H, 2 x OH), 4.21 (q, *J* = 7.1 Hz, 2H, CH_2), 6.35 (d, *J* = 16.0 Hz, 1H, Ar-CH=*CH*), 7.18 (d, *J* = 2.0 Hz, 1H, Ar-H), 7.39 (d, *J* = 2.0 Hz, 1H, Ar-H), 7.52 (d, *J* = 15.5 Hz, 1H, Ar-CH).

^{13}C-NMR (500 MHz, $(CD_3)_2CO$) δ: 14.1 (CH_3), 60.2 (CH_2), 109.7 (Ar-C-Br), 114.7 (Ar-CH), 117.0 (Ar-CH=*CH*), 124.6 (Ar-CH), 127.9 (Ar-C-OH), 143.5 (Ar-*CH*=CH), 145.6 (Ar-C-OH), 166.6 (C=O).

HRMS-ES (*m/z*): $[M - H]^-$ calcd for $C_{11}H_{10}O_4Br$, 284.9762; found, 284.9775.

5.2.1.15. *(E)-ethyl 3-(3,4-bis(2-(benzyloxycarbonylamino)ethoxy)-5-bromophenyl)-acrylate* **28**



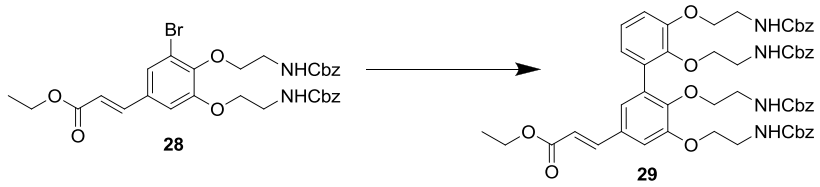
To a stirring solution of (*E*)-ethyl 3-(3-bromo-4,5-dihydroxyphenyl)acrylate **27** (1.21g, 4.23 mmol) in acetone (12 mL) was added Cs_2CO_3 (2.06 g, 6.35 mmol) and the mixture stirred at 65°C under an argon atmosphere until a yellow salt was formed, at which point mesylate **3** was added (3.46 g, 12.7 mmol) and the mixture stirred for 17 h. The solvent was removed under vacuum and 1M HCl (50 mL) added. The product was extracted with ethyl acetate (3 x 20 mL), washed with brine, dried over $MgSO_4$, filtered and concentrated under vacuum. The product was adsorbed on silica and purified by flash column chromatography using a gradient of 0 to 40% ethyl acetate in cyclohexane to yield the title product which was recrystallised from a mixture of methanol and water (3:1) (2.62 g, 96%).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.36 (t, *J* = 7.1 Hz, 3H, CH_3), 3.48-3.64 (m, 4H, 2 x CH_2), 3.98-4.20 (m, 4H, 2 x CH_2), 4.29 (q, *J* = 7.1 Hz, 2H, CH_2CH_3), 5.03-5.24 (m, 4H, 2 x Ph-CH_2), 6.36 (d, *J* = 15.9 Hz, 1H, Ar-CH=*CH*), 7.00 (s, 1H, ArH), 7.26-7.44 (m, 11H, 11 x ArH), 7.53 (d, *J* = 16.0 Hz, 1H, Ar-*CH*=CH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 14.7 (CH_3), 40.7 (Ar-O-CH_2CH_2), 61.0 (CH_2CH_3), 67.3 (O-CH_2-Ar), 68.7 (Ar-O-CH_2), 112.05 (ArC), 119.6 (Ar-CH=*CH*), 125.8 (ArC), 127.2 (ArCH), 128.3 (ArCH), 128.9 (ArCH), 136.7 (ArC), 142.8 (Ar-*CH*=CH), 153.0 (ArC), 156.8 (ArC), 166.9 (C=O).

HRMS-ES (*m/z*): $[M + Na]^+$ calcd $C_{31}H_{33}N_2O_8NaBr$, 663.1318; found, 663.1301.

5.2.1.16. *(E)-ethyl 3-(2',3',5,6-tetrakis(2-(benzyloxycarbonylamino)ethoxy)biphenyl-3-yl)acrylate* **29**



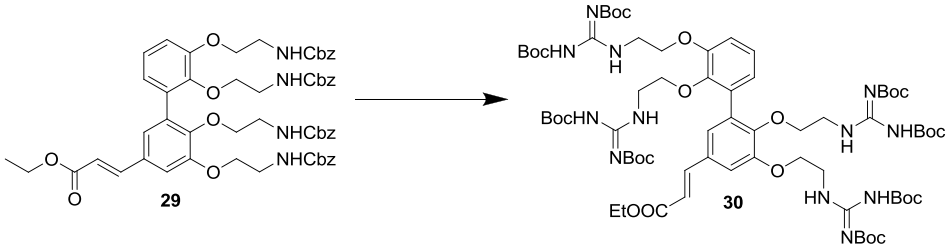
To a degassed solution of potassium 2,3-bis(2-benzyloxycarbonylamino)ethoxy)-phenyltrifluoroborate **18** (1.24 g, 2.18 mmol) in isopropanol/water (2:1) (21 ml) was added $PdCl_2dppf.CH_2Cl_2$ (63 mg, 7%) and potassium carbonate (309 mg, 2.18 mmol) and the mixture stirred for 2 min. (*E*)-ethyl-3-(3,4-bis(2-(benzyloxycarbonylamino)-ethoxy)-5-bromophenyl)acrylate **28** (700 mg, 1.09 mmol) was added and the mixture irradiated in a microwave at 80 °C while stirring for 10 min. The mixture was cooled to room temperature, the catalyst removed by filtration and the mixture was poured onto 1M HCl (20 mL). The product was extracted with diethyl ether, dried over $MgSO_4$, filtered and concentrated under vacuum. The mixture was adsorbed on silica and purified by flash column chromatography to yield the title product as a colourless oil (952 mg, 85%).

^{1}H-NMR (500 MHz, $(CD_3)_2CO$) δ: 1.29 (t, *J* = 7.1 Hz, 3H, CH_3), 3.17-3.29 (m, 4H, 2 x CH_2), 3.57-3.66 (m, 4H, 2x CH_2), 3.90-4.04 (m, 4H, 2 x CH_2), 4.08-4.16 (m, 4H, 2 x CH_2), 4.22 (q, *J* = 7.1 Hz, 2H, CH_2CH_3), 4.98-5.16 (m, 8H, 4 x O-CH_2-Ar), 6.56 (d, *J* = 15.9 Hz, 1H, Ar-CH=*CH*), 7.65 (d, *J* = 16.0 Hz, 1H, Ar-*CH*=CH), 7.23-7.45 (m, 25H, 25 x ArH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 14.4 (CH_3), 41.2 (CH_2NH), 60.6 (CH_2CH_3), 65.9 (O-CH_2-Ar), 66.8 (Ar-O-CH_2CH_2), 112.1 (CH), 113.7 (CH), 118.2 (CH), 123.3 (CH), 124.0 (CH), 124.4 (CH), 128.6 (ArCH), 136.5 (ArC), 143.7 (CH), 156.4 (NH*C*OO), 166.9 (CH=CH*C*=O).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_57H_{61}N_4O_{14}$, 1025.4184; found, 1025.4120.

5.2.1.17. *(2E)-ethyl 3-(2',3',5,6-tetrakis(2-(2,3-bis(tert-butoxycarbonyl)guanidino)ethoxy)biphenyl-3-yl)acrylate* **30**



To a stirring solution of (*E*)-ethyl-3-(2',3',5,6-tetrakis(2-(benzyloxycarbonylamino)-ethoxy)bi-phenyl-3-yl)acrylate **29** (930 mg, 0.91 mmol) in DCM (20 mL) was added 30% HBr in acetic acid (7 mL) dropwise. The mixture was stirred for 3 h and then diluted with DCM (20 mL) and the amine extracted with water (3 x 10 mL) and dried under vacuum for several hours. The resulting orange solid was suspended in DCM (30 mL), DIEA was added (1.58 mL, 9.01 mmol) and the mixture stirred for 10 min. *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine (1.27 g, 4.08 mmol) was added and the mixture stirred for 17 h at room temperature under a nitrogen atmosphere. The reaction mixture was diluted with DCM (40 mL), washed with 10% citric acid (20 mL), sat. $NaHCO_3$ and brine, and the organic layer dried over $MgSO_4$, filtered and concentrated under vacuum. The product was purified by flash column chromatography with a 0 to 30% ethyl acetate gradient in cyclohexane to yield the title compound as a colourless oil (725 mg, 55%).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.32 (t, *J* = 7.1 Hz, 3H, $CH_2\mathit{CH_3}$), 1.40-1.48 (m, 72H, 24 x Boc-CH_3), 3.44-3.50 (m, 4H, 2 x CH_2), 3.84-3.99 (m, 8H, 4 x CH_2), 4.15-4.22 (m, 4H, 2 x CH_2), 4.24 (q, *J* = 7.2 Hz, 2H, $\mathit{CH_2}CH_3$), 6.32 (d, *J* = 15.9 Hz, 1H, Ar-CH=*CH*), 6.88-7.13 (m, 5H, ArH), 7.58 (d, *J* = 15.9 Hz, 1H, Ar-*CH*=CH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 14.4 ($CH_2\mathit{CH_3}$), 28.2 (24 x Boc-CH_3), 40.1 (CH_2), 41.5 (CH_2), 60.4 ($\mathit{CH_2}CH_3$), 67.4 (CH_2), 70.9 (CH_2), 78.9 (Boc-C), 79.3 (Boc-C), 111.7 (ArCH), 113.2 (ArCH), 117.8 (Ar-CH=*CH*), 123.6 (ArCH), 124.0 (ArCH), 124.6 (ArCH), 310.2 (ArC), 132.2 (ArC), 133.2 (ArC), 144.1 (Ar-*CH*=CH), 145.3 (ArC), 147.4 (ArC), 151.8 (Boc-C=O), 153.0 (Boc-C=O), 156.2 (C=N), 167.0 ($CH_3CH_2\mathit{C}OO$).

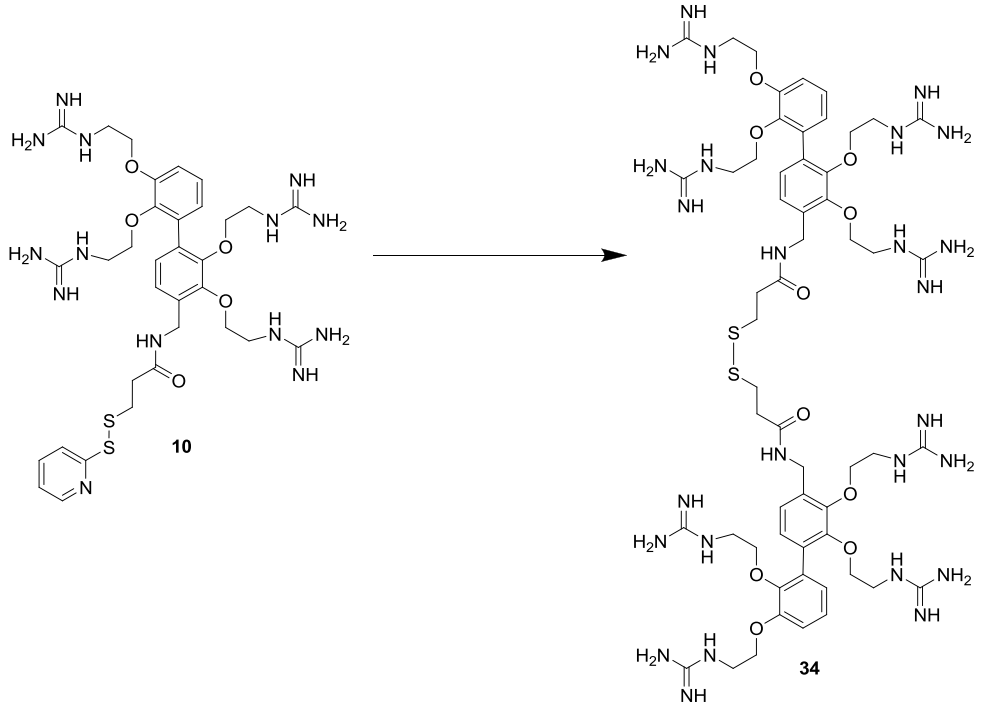
HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{69}H_{109}N_{12}O_{22}$, 1457.7779; found, 1457.7532.

5.2.1.18. *Ester hydrolysis*

Method 1: (2E)-ethyl 3-(2',3',5,6-tetrakis(2-(2,3-bis(tert-butoxycarbonyl)guanidino)-ethoxy)biphenyl-3-yl)acrylate **30** (100 mg, 0.07 mmol) was dissolved in the required solvent (Table 7) (1 mL) and NaOH was added (either 2-6 eq, or 1 mL of a 1 M solution). The reaction was stirred at room temperature or 50°C for 4-5 h and the reaction progress monitored by LC-MS.

Method 2: (2E)-ethyl 3-(2',3',5,6-tetrakis(2-(2,3-bis(tert-butoxycarbonyl)guanidino)-ethoxy)biphenyl-3-yl)acrylate **30** (100 mg, 0.07 mmol) was dissolved in THF/water (1:1) and cooled to 0°C. Lithium hydroxide (8.4 mg, 0.35 mmol) was added and the mixture warmed to room temperature and stirred for 24 h. The progress of the reaction was monitored by LC-MS.

5.2.1.19. *3,3'-disulfanediylbis(N-((2,2',3,3'-tetrakis(2-guanidinoethoxy)biphenyl-4-yl)methyl)propanamide)* **34**



2,2’,2’’,2’’’-(4-((3-(pyridine-2-yldsulfanyl)propanamido)methyl)biphenyl-2,2',3,3'-tetrayl)-tetrakis(oxy)tetrakis(ethane-2,1-diyl)tetrakis(azanediyl)tetrakis(amino-methaniminium) **10** (4 mg, 3.22 μmol) was dissolved in methanol/water (1/19) (300 μL) and dithiothreitol (9.9 mg, 64.4 μl) was added. The reaction was stirred for 30 min and monitored by LC-MS. On completion, the thiol product **33** was purified by preparatory LC/MS on a C18 preparatory column using a mobile phase of water/acetonitrile (95/5). The solvent was removed under vacuum and the product dissolved in a solution of ammonium acetate (100 μL, 7 g/L) and stirred for 72 h. The product was washed with diethyl ether and lyophilised to yield the title compound **34** as a white solid (2.2 mg, 55%).

LRMS-MALDI (*m/z*): $[M + H]^+$ calcd for $C_{56}H_{88}N_{26}O_{10}S_2$, 1349.66; found, 1350.35.

5.3. Chapter 3 Methods

5.3.1. Molecular Dynamics

Molecular dynamics simulations were run for compounds **43**, **44**, **45** and **46** in the YASARA[161] software using the AMBER99 force field[162] with default parameters. The starting structures were contained in a simulation cell which extended 5 Å larger than the SMoC structure along each axis in vacuo, the default pH set to 7.0 and the structures were minimised under the AMBER99 force field starting at 298 K for 1 ps. The simulation substep time was set to 1 fs and the intramolecular forces were recalculated every 2 substeps and the temperature controlled by velocity rescaling. A 5 ns simulation was performed for each structure, after which snapshots were taken and analysed in the DS Visualizer software. No atoms were kept frozen during the simulations.

5.3.2. NMR study

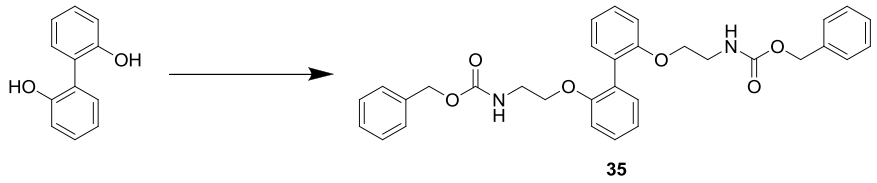
The SMoC compounds were dissolved in 0.5 mL deuterium oxide to a final concentration of around 10 mg/ml. 100 mM NaOH was added in small aliquots (5-30 µl) and the pH of the solution measured in the NMR tube using a micro electrode attached to a pH meter. NMR spectra were recorded using a 400 MHz Bruker spectrometer at 20 °C.

5.3.3. ITC

6G-SMoC at a concentration of 250 µM was loaded into the syringe of a Microcal VP-ITC calorimeter (450 µl), and the cell (1.8 ml) filled with human GAPDH siRNA (Dharmacon RNAi Technologies) at a concentration of 3.5 µM. A total of 37 injections of 8 µl each were made at 4 min intervals to ensure total saturation of the siRNA. The integrated heats evolved for each injection were plotted against the molar ratio of the mixture in the cell, and a binding curve fitted using non-linear least squares analysis of the data. The thermodynamic parameters of the binding were calculated using a one-site (or *n* identical sites) binding site model in the Origin 6.0 software (Microcal).

5.3.4. Syntheses

5.3.4.1. *Benzyl 2,2'-(biphenyl-2,2'-diylbis(oxy))bis(ethane-2,1-diyl)dicarbamate* **35**



35

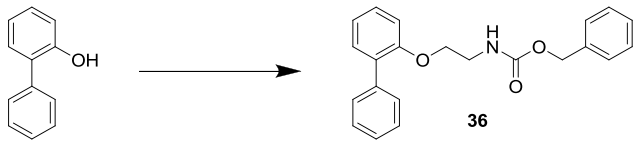
2,2'-dihydroxybiphenyl (100 mg, 0.54 mmol) was stirred in acetone (1.5 mL) with caesium carbonate (228 mg, 0.70 mmol) at 65 °C under nitrogen. When a white salt appeared, 2-(benzyloxycarbonylamino)ethyl methanesulfonate (381 mg, 1.40 mmol) was added and the reaction stirred overnight. The mixture was concentrated under vacuum and adsorbed onto silica gel for purification by flash column chromatography using ethyl acetate/cyclohexane (1/2) as eluent to yield the title compound (100 mg, 35% yield) as a viscous colourless oil.

^{1}H-NMR (500 MHz, CD_3OD) δ: 3.30 (t, *J* = 5.6 Hz, 4H, 2 x *CH*$_2$NH), 3.93 (t, *J* = 5.6 Hz, 4H, 2 x *CH*$_2$CH$_2$NH), 5.00 (s, 4H, 2 x O*CH*$_2$Ph), 6.92-7.31 (m, 18H, 18 x ArH).

^{13}C-NMR (500 MHz, CD_3OD) δ: 40.5 (2 x NH*CH*$_2$) 67.6 (O*CH*$_2$Ph), 68.7 (*CH*$_2$CH$_2$NH), 122.3 (ArCH), 128.2 (ArCH), 129.0 (ArCH), 129.5 (ArCH), 129.8 (ArC), 132.3 (ArC), 157.4 (C=O), 159.3 (ArC-O).

LRMS-ES (*m/z*): $[M + Na]^+$ calcd for $C_{32}H_{32}N_2O_6Na$, 563.23; found, 564.20.

5.3.4.2. *Benzyl 2-(biphenyl-2-yloxy)ethylcarbamate* **36**



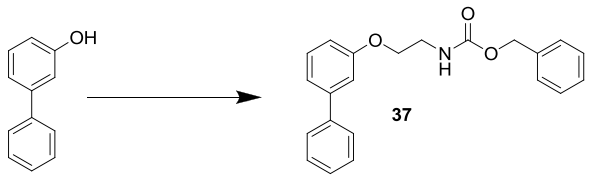
2-Hydroxybiphenyl (1 g, 5.88 mmol) was dissolved in acetone (15 mL) with Cs_2CO_3 (2.49 g, 7.64 mmol) at 65°C under an argon atmosphere for 10 min. 2-(benzyloxycarbonylamino)ethyl methanesulfonate was added (4.17 g, 15.29 mmol) and the mixture was stirred for 17 h. When the reaction was confirmed to be complete by tlc, the solvent was removed under vacuum and water was added (30 mL). The product was extracted with ethyl acetate (3 x 25 mL) and the organic layers combined, dried over $MgSO_4$, filtered and concentrated under vacuum. The mixture was purified by flash chromatography using an ethyl acetate gradient of 0-30% in cyclohexane. The product was obtained as a colourless oil (1.9 g, 93%).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.52 (q, *J* = 5.4 Hz, 2H, *CH*$_2$NH), 4.04 (t, *J* = 5.0 Hz, 2H, *CH*$_2$CH$_2$NH), 5.13 (s, 2H, O*CH*$_2$Ph), 7.01 (d, *J* = 8.2 Hz, 1H, ArH), 7.11 (td, *J* = 1.08, 7.47 Hz, 1H, ArH), 7.29-7.49 (m, 10H, 10 x ArH), 7.55 (m, 2H, 2 x ArH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 40.7 (*CH*$_2$NH), 66.8 (O*CH*$_2$Ph), 68.0 (*CH*$_2$CH$_2$NH), 113.6 (ArCH), 121.9 (ArCH), 128.1 (ArCH), 128.2 (ArCH), 128.3 (ArCH), 128.6 (ArCH), 129.6 (ArCH), 131.0 (ArCH), 131.6 (ArC), 136.7 (ArC), 138.5 (ArC), 155.5 (C=O), 156.5 (ArC-O).

HRMS-ES (*m/z*): $[M + Na]^+$ calcd for $C_{22}H_{21}NO_3Na$, 370.1419; found, 370.1408.

5.3.4.3. *Benzyl 2-(biphenyl-3-yloxy)ethylcarbamate* **37**



37

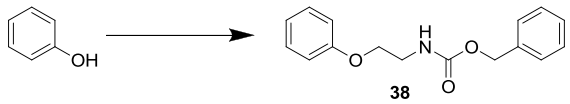
3-Hydroxybiphenyl (1 g, 5.88 mmol) was dissolved in acetone (15 mL) with Cs_2CO_3 (2.49 g, 7.64 mmol) at 65 °C under an argon atmosphere for 10 min. 2-(Benzyloxycarbonylamino)ethyl methanesulfonate was added (2.40 g, 8.82 mmol) and the mixture was stirred for 17 h. When the reaction was confirmed to be complete by tlc, the solvent was removed under vacuum and water was added (30 mL). The product was extracted with ethyl acetate (3 x 25 mL) and the organic layers combined, dried over $MgSO_4$, filtered and concentrated under vacuum. The mixture was purified by flash chromatography using an ethyl acetate gradient of 0-30% in cyclohexane. The product was obtained as a colourless oil (2.04 g, 100% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.57-3.68 (m, 2H, *CH*$_2$NH), 4.10 (t, *J* = 5.0 Hz, 2H, *CH*$_2$CH$_2$NH), 5.14 (s, 2H, O*CH*$_2$Ph), 6.87-7.60 (m, 14H, 14 x ArH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 40.8 (*CH*$_2$NH), 60.5 (O*CH*$_2$Ph), 67.0 (*CH*$_2$CH$_2$NH), 113.3 (ArCH), 113.6 (ArCH), 120.2 (ArCH), 127.1 (ArCH), 127.3 (ArCH), 127.6 (ArCH), 128.3 (ArCH), 128.6 (ArCH), 128.9 (ArCH), 129.0 (ArCH), 136.5 (ArCH), 141.0 (ArC), 141.2 (ArC), 143.0 (ArC), 156.6 (C=O), 159.0 (ArC-O).

HRMS-ES (*m/z*): $[M + Na]^+$ calcd for $C_{22}H_{21}NO_3Na$, 370.1419; found, 370.1415.

5.3.4.4. *Benzyl 2-phenoxyethylcarbamate* **38**



38

Phenol (1 g, 10.6 mmol) was dissolved in acetone (15 mL) with Cs_2CO_3 (4.5 g, 15.9 mmol) at 65 °C under an argon atmosphere for 10 min. 2-(benzyloxycarbonyl-amino)ethyl methanesulfonate was added (4.35 g, 15.9 mmol) and the mixture was stirred for 17 h. When the reaction was confirmed to be complete by tlc, the solvent was removed under vacuum and water was added (30 mL). The product was extracted with ethyl acetate (3 x 25 mL) and the organic layers combined, dried over $MgSO_4$, filtered and concentrated under vacuum. The mixture was purified by flash chromatography using an ethyl acetate gradient of 0-30% in cyclohexane. The product was obtained as a colourless oil (1.91 g, 70%).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.47-3.71 (m, 2H, *C*H_2NH), 4.03 (t, *J* = 5.0 Hz, 2H, *C*H_2CH_2NH), 5.15 (s, 2H, O*C*H_2Ph), 6.90-7.40 (m, 10H, 10 x ArH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 40.7 (*C*H_2NH), 65.3 (O*C*H_2Ph), 66.9 (*C*H_2CH_2NH), 114.6 (Ar*C*H), 121.3 (Ar*C*H), 128.3 (Ar*C*H), 128.7 (Ar*C*H), 129.7 (Ar*C*H), 136.6 (Ar*C*), 156.7 (*C*=O), 158.6 (Ar*C*-O).

HRMS-EI (*m/z*): $[M]^+$ calcd for $C_{16}H_{17}O_3N$, 271.12029; found, 271.12069.

5.3.4.5. *Tert-butyl (tert-butoxycarbonylamino)(2-phenoxyethylamino)methylenecarbamate* **39**

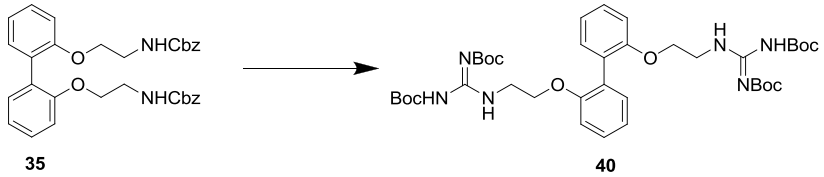
Benzyl 2-phenoxyethylcarbamate **38** (1.91 g, 7.05 mmol) was dissolved in DCM (20 mL) and 33% HBr in acetic acid (10 mL) was added dropwise while stirring. The reaction was stirred for 2.5 h before tlc confirmed the reaction to be complete. The solvent was evaporated and the product dried under vacuum for several hours. The resulting orange solid was dissolved in acetonitrile (50 mL), to which was added DIEA (12.3 mL, 70.5 mmol) and *N*,*N*′-Di-Boc-1*H*-pyrazole-1-carboxamidine (2.3 g, 7.4 mmol). The reaction was stirred for 48 h under an argon atmosphere. The solvent was evaporated and water added (50 mL). The product was extracted with ethyl acetate (3 x 20 mL) and the organic extracts were combined and washed with brine (50 mL), dried over $MgSO_4$, filtered and concentrated under vacuum. The product was purified by flash chromatography using an ethyl acetate gradient of 0-50% in cyclohexane. The product was dried under vacuum to form a white solid (1.21 g, 45% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.49 (s, 9H, 3 x Boc-*CH*$_3$), 1.51 (s, 9H, 3 x Boc-*CH*$_3$), 3.81-3.87 (m, 2H, *CH*$_2$NH), 4.09 (t, *J* = 5.4 Hz, 2H, *CH*$_2$CH$_2$NH), 6.93-6.96 (m, 3H, 3 x ArH), 7.21-7.30 (m, 2H, 2 x ArH), 8.75 (br, 1H, NH), 11.48 (br, 1H, NH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 28.4 (Boc-*CH*$_3$), 40.3 (*CH*$_2$NH), 66.3 (*CH*$_2$CH$_2$NH), 79.4 (*C*(CH$_3$)$_3$), 83.2 (*C*(CH$_3$)$_3$), 114.8 (Ar*C*H), 121.2 (Ar*C*H). 129.6 (Ar*C*H), 153.1 (*C*=O), 156.4 (*C*=O), 158.6 (Ar*C*-O), 163.6 (*C*=N).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{19}H_{30}N_3O_5$, 380.2185; found, 380.2174.

5.3.4.6. *Tert-butyl (2,2'-(biphenyl-2,2'-diylbis(oxy))bis(ethane-2,1-diyl))bis(azanediyl) bis((tert-butoxycarbonylamino)methan-1-yl-1-ylidene)dicarbamate* **40**



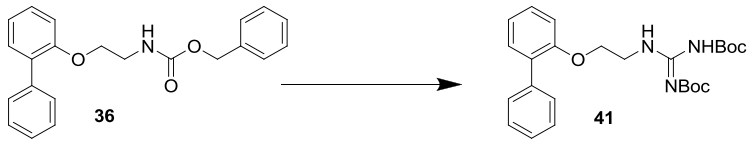
35 **40**

Benzyl 2,2'-(biphenyl-2,2'-diylbis(oxy))bis(ethane-2,1-diyl)dicarbamate **35** (80 mg, 0.15 mmol) was dissolved in DCM (10 mL) and 30% HBr in acetic acid (2 mL) was added dropwise while stirring and the reaction was stirred for 3.5 h at rt. Water (5 mL) was added and the layers separated. The aqueous layer was washed with DCM (2 x 5 ml) and the water removed under vacuum. The resulting orange salt was dried for several hours under vacuum and used without further purification. The solid was dissolved in acetonitrile (5 mL) with *N,N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine (214 mg, 0.7 mmol) and DIEA was added dropwise (0.6 mL, 3.5 mmol). The mixture was stirred at room temperature under argon for 24 h. The reaction was monitored by LC-MS until all the starting material was consumed and only the di-guanidinylated product was observed. The solvent was removed under vacuum and the product purified by flash chromatography (0-20% ethyl acetate gradient in cyclohexane). The title product was obtained as a white solid (58 mg, 51% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.48 (s, 18H, 6 x Boc-*CH*$_3$), 1.49 (s, 18H, 6 x Boc-*CH*$_3$) 3.65-3.70 (m, 4H, 2 x *CH*$_2$NH), 4.02 (t, *J* = 5.04 Hz, 4H, 2 x *CH*$_2$CH$_2$NH), 6.90-7.09 (m, 4H, 4 x ArH), 7.18-7.40 (m, 4H, 4 x ArH).

HRMS-FAB (*m/z*): $[M + Na]^+$ calcd for $C_{38}H_{56}N_6O_{10}Na$, 779.39554; found, 779.39438.

5.3.4.7. *Tert-butyl (2-(biphenyl-2-yloxy)ethylamino)(tert-butoxycarbonylamino)methylenecarbamate* **41**



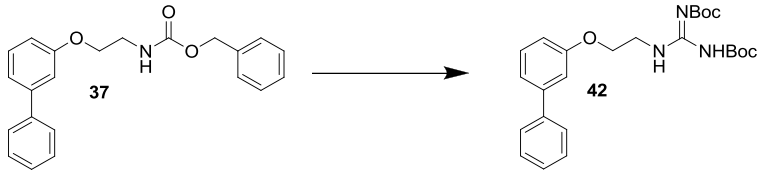
Benzyl 2-(biphenyl-2-yloxy)ethylcarbamate **36** (1.9 g, 5.46 mmol) was dissolved in DCM (20 mL) and 33% HBr in acetic acid (10 mL) was added dropwise while stirring. The reaction was stirred for 2.5 h before tlc confirmed the reaction to be complete. The solvent was evaporated and the product dried under vacuum for several hours. The resulting orange solid was dissolved in acetonitrile (50 mL), to which was added DIEA (9 mL, 54.6 mmol) and *N*,*N*′-Di-Boc-1*H*-pyrazole-1-carboxamidine (1.7 g, 5.46 mmol). The reaction was stirred for 48 h under an argon atmosphere. The solvent was evaporated and water added (50 mL). The product was extracted with ethyl acetate (3 x 20 mL) and the organic extracts were combined and washed with brine (50 mL), dried over $MgSO_4$, filtered and concentrated under vacuum. The product was purified by flash chromatography using an ethyl acetate gradient of 0-50% in cyclohexane. The product was dried under vacuum to form a colourless oil (0.64 g, 26% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 2.74 (s, 18H, 6 x Boc-*CH*$_3$), 5.02 (t, *J* = 5.0 Hz, 2H, 2 x *CH*$_2$NH), 5.30 (t, *J* = 4.7 Hz, 2H, 2 x *CH*$_2$CH$_2$NH), 8.19-8.80 (m, 9H, 9 x ArH), 9.95 (br, 1H, NH), 12.72 (br, 1H, NH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 29.4 (Boc-*CH*$_3$), 29.6 (Boc-*CH*$_3$), 41.5 (*CH*$_2$NH), 68.1 (*CH*$_2$CH$_2$NH), 84.5 (*C*(CH$_3$)$_3$), 113.6 (ArCH), 122.7 (ArCH), 128.2 (ArCH), 129.3 (ArCH), 129.6 (ArCH), 129.9 (ArCH), 130.9 (ArCH), 132.3 (ArC), 139.6 (ArC), 154.3 (C=O), 156.6 (ArC-O), 157.6 (C=N).

HRMS-CI (*m/z*): $[M + H]^+$ calcd for $C_{25}H_{34}N_3O_5$, 456.24985; found, 456.24990.

5.3.4.8. *(Z)-tert-butyl (2-(biphenyl-3-yloxy)ethylamino)(tert-butoxycarbonylamino)methylenecarbamate* **42**



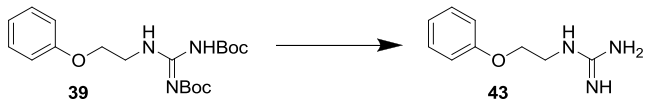
Benzyl 2-(biphenyl-3-yloxy)ethylcarbamate **37** (2 g, 5.75 mmol) was dissolved in DCM (20 mL) and 33% HBr in acetic acid (10 mL) was added dropwise while stirring. The reaction was stirred for 2.5 h before tlc confirmed the reaction to be complete. The solvent was evaporated and the product dried under vacuum for several hours. The resulting orange solid was dissolved in acetonitrile (50 mL), to which was added DIEA (9 mL, 54.6 mmol) and *N*,*N*′-Di-Boc-1*H*-pyrazole-1-carboxamidine (1.7 g, 5.46 mmol). The reaction was stirred for 48 h under an argon atmosphere. The solvent was evaporated and water added (50 mL). The product was extracted with ethyl acetate (3 x 20 mL) and the organic extracts were combined and washed with brine (50 mL), dried over $MgSO_4$, filtered and concentrated under vacuum. The product was purified by flash chromatography using an ethyl acetate gradient of 0-50% in cyclohexane. The product was dried under vacuum to form white solid (1.05 g, 40% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.49 (s, 9H, 3 x Boc *CH*$_3$), 1.51 (s, 9H, 3 x Boc-*CH*$_3$), 3.85-3.93 (m, 2H, *CH*$_2$NH), 4.16 (t, *J* = 5.4 Hz, 2H, *CH*$_2$CH$_2$NH), 6.92-7.59 (m, 9H, 9 x ArH), 8.81 (br, 1H, NH), 11.50 (br, 1H, NH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 28.4 (Boc-*CH*$_3$), 40.3 (*CH*$_2$NH), 66.3 (*CH*$_2$CH$_2$NH), 79.4 (*C*(CH$_3$)$_3$), 83.2 (*C*(CH$_3$)$_3$), 114.8 (ArCH), 121.2 (ArCH), 129.6 (ArCH), 153.1 (C=O), 156.4 (C=O), 159.0 (ArC-O), 163.6 (C=N).

HRMS-CI (*m/z*): $[M + H]^+$ calcd for $C_{25}H_{34}N_3O_5$, 456.24985; found, 456.25037.

5.3.4.9. *1-(2-phenoxyethyl)guanidine* **43**



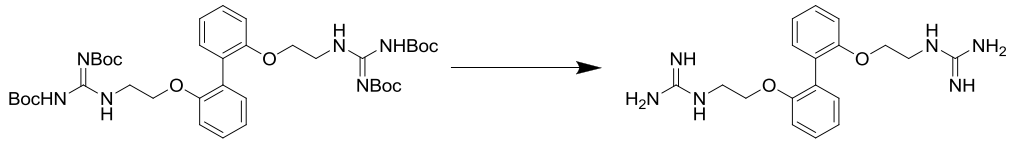
39 43

Tert-butyl-(tert-butoxycarbonylamino)(2-phenoxyethylamino)methylenecarbamate **39** (179.95 mg, 0.47 mmol) was dissolved in a mixture of TFA/TIPS/water (95:2.5:2.5, 15 ml) and stirred at room temperature for 4 h. The solvent was removed under vacuum and the product washed with ether (115.9 mg, 100% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.44-3.49 (m, 2H, *CH*$_2$NH), 4.03-4.09 (s, 2H, *CH*$_2$CH$_2$NH), 6.82-6.98 (m, 3H, 3 x ArH), 7.12-7.40 (m, 2H, 2 x ArH).

HRMS-CI (*m/z*): $[M + H]^+$ calcd for $C_9H_{14}N_3O$, 180.11369; found, 180.11381.

5.3.4.10. *1,1'-(2,2'-(biphenyl-2,2'-diylbis(oxy))bis(ethane-2,1-diyl))diguanidine* **44**



40 44

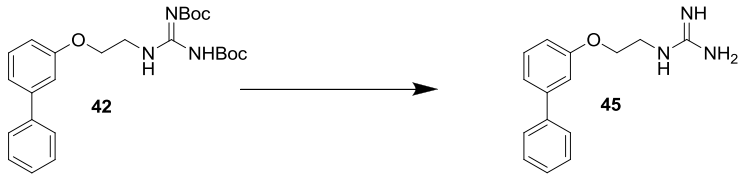
Tert-butyl-(2,2'-(biphenyl-2,2'-diylbis(oxy))bis(ethane-2,1-diyl))bis(azanediyl)bis-((tertbutoxycarbonylamino)methan-1-yl-1-ylidene)dicarbamate **40** (258.5 mg, 0.34 mmol) was dissolved in TFA/TIPS/H_2O (95:2.5:2.5, 20 mL) and stirred at room temperature for 3 h. The solvent was removed under vacuum and the product washed with ether (198 mg, 100% yield).

^{1}H-NMR (500MHz, CD_3OD) δ: 3.43 (t, *J* = 4.7 Hz, 4H, 2 x *CH*$_2$NH), 4.05 (t, *J* = 4.7 Hz, 4H, 2 x *CH*$_2$CH$_3$NH), 7.02-7.09 (m, 4H, 4 x ArH), 7.16-7.21 (m, 2H, 2 x ArH), 7.31-7.39 (m, 2H, 2 x ArH).

^{13}C-NMR (500 MHz, CD_3OD) δ: 42.5 (*CH*$_2$NH), 69.0 (*CH*$_2$CH$_2$NH), 113.8 (ArCH), 122.6 (ArCH), 129.4 (ArCH), 130.2 (ArCH), 132.5 (ArC), 157.1 (ArC-O).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{18}H_{25}N_6O_2$, 357.2039; found, 357.2050.

5.3.4.11. *1-(2-(biphenyl-3-yloxy)ethyl)guanidine* **45**



(Z)-*tert*-butyl(2-(biphenyl-3-yloxy)ethylamino)(tertbutoxycarbonylamino)methyl-enecarbamate **42** (110 mg, 0.24 mmol) was dissolved in a mixture of TFA/TIPS/water (95:2.5:2.5, 15 ml) and stirred at room temperature for 4 h. The solvent was removed under vacuum and the product washed with ether (95.1 mg, 100%).

^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.83 (s, 2H, *CH*$_2$NH), 6.83 (s, 2H, *CH*$_2$CH$_2$NH), 9.31-9.81 (m, 9H, 9 x ArH), 10.39 (br, 2H 2 x NH).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{15}H_{18}N_3O$, 256.1450; found, 256.1441.

5.3.4.12. *1-(2-(biphenyl-2-yloxy)ethyl)guanidine* **46**

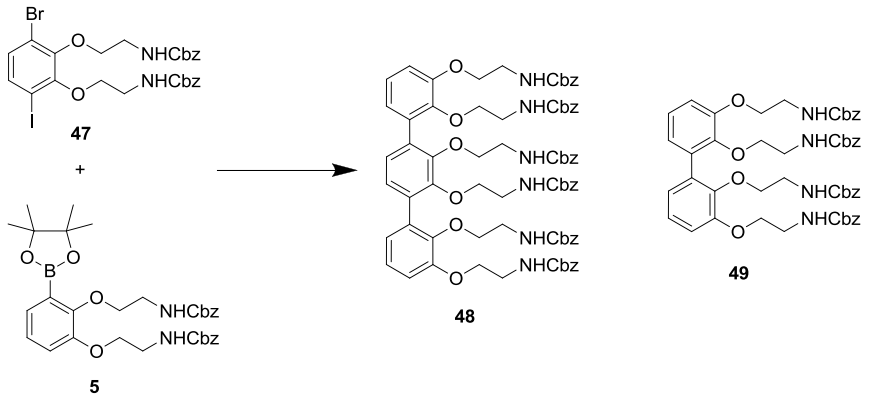
Tert-butyl (2-(biphenyl-2-yloxy)ethylamino)(tertbutoxycarbonylamino)methylene-carbamate **41** (100 mg, 0.23 mmol) was dissolved in a mixture of TFA/TIPS/water (95:2.5:2.5, 15 ml) and stirred at room temperature for 3 h. The solvent was removed under vacuum and the product washed with ether (84.8 mg, 100% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.45 (br, 2H, CH_2NH), 4.06 (t, J = 3.5 Hz, 2H, CH_2CH_2NH), 7.07-7.41 (m, 9H, 9 x ArH), 11.87 (br, 2H, 2 x NH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 42.1 (CH_2NH), 69.1 (CH_2CH_2NH), 112.6 (ArCH), 122.8 (ArCH), 127.9 (ArCH), 128.0 (ArCH), 129.2 (ArCH), 130.0 (ArC), 138.4 (ArC-O), 158.7 (C=N).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{15}H_{18}N_3O$, 256.1450; found, 256.1444.

5.3.4.13. *6G-SMoC-NHCbz* **48**



A degassed mixture of dibenzyl 2,2'-(3-bromo-6-iodo-1,2-phenylene) bis(oxy)bis(ethane-2,1-diyl)dicarbamate **47** (1 g, 1.5 mmol), dibenzyl 2,2'-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1,2-phenylene)bis(oxy)bis(ethane-2,1-diyl)di-carbamate **5** (1.77 g, 3 mmol), $PdCl_2$dppf. CH_2Cl_2 (61 mg, 0.075 mmol), potassium phosphate (0.69 g, 3 mmol), toluene (10 mL) and water (1 mL) was heated at 100°C for 3 h. The reaction mixture was diluted in EtOAc (25 mL) and water (25 mL) was added. The layers were separated and the aqueous layer was extracted with EtOAc (3 x 25 mL). The combined organic layers were dried over $MgSO_4$, filtered and concentrated under vacuum. The residue was then purified by flash chromatography using cyclohexane/EtOAc (1/1) as eluent to afford **48** (0.21 g, 10% yield) as a yellow oil and benzyl 2,2',2",2"'-(4-bromobiphenyl-2,2',3,3'-tetrayl)tetrakis(oxy)tetrakis-(ethane-2,1-diyl)tetracarbamate **49** (1.09 g, 1.08 mmol). A degassed mixture of **49** (1.09 g, 1.08 mmol), dibenzyl 2,2'-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1,2-phenyl-ene)bis(oxy)bis(ethane-2,1-diyl)dicarbamate (0.94 g, 1.6 mmol), $PdCl_2$dppf.CH_2Cl_2 (40 mg, 0.05 mmol), potassium phosphate (0.46 g, 2 mmol), toluene (7.5 mL) and water (0.75 mL) was heated at 100°C overnight. The reaction mixture was diluted in EtOAc (25 mL) and water (25 mL) was added. The layers were separated and the aqueous layer was extracted with EtOAc (3 x 25 mL). The combined organic layers were dried over $MgSO_4$, filtered and concentrated under vacuum. The residue was then purified by flash

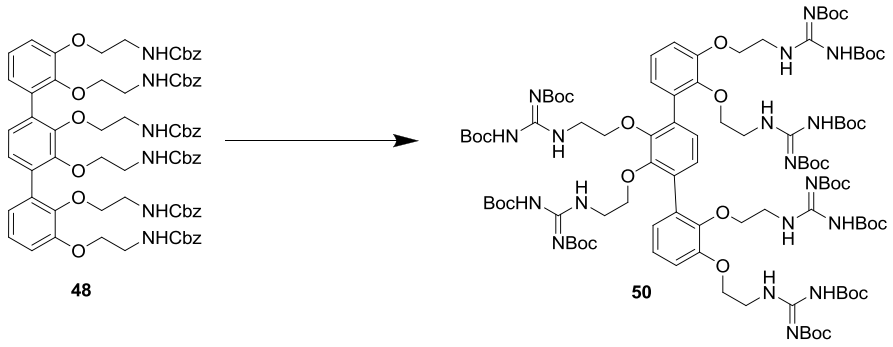
chromatography using cyclohexane/EtOAc (1/1) as eluent to afford **48** (0.88 g, 59% yield) as a yellow oil. The global yield was 48%.

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.02-3.33 (m, 8H, 4 x CH_2), 3.38-3.64 (m, 4H, 2 x CH_2), 3.67-4.05 (m, 12H, 6 x CH_2), 4.92-5.21 (m, 12H, 6 x O*C*H_2Ph), 6.66-7.63 (m, 38H, 38 x ArH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 40.7 (*C*H_2NH), 41.0 (*C*H_2NH), 41.3 (*C*H_2NH), 66.7 (O*C*H_2Ph), 68.0 (*C*$H_2$$CH_2$NH), 72.4 (*C*$H_2$$CH_2$NH), 113.4 (ArCH), 123.4 (ArCH), 124.3 (ArCH), 126.1 (ArCH), 128.2 (ArCH), 128.5 (ArCH), 128.7 (ArCH), 132.4 (ArC), 136.4 (ArC), 136.7 (ArC), 151.6 (ArC-O), 156.3 (ArC-O), 156.5 (ArC-O), 171.2 (C=O).

HRMS-ES (*m/z*): $(M + Na)^+$ calcd for $C_{78}H_{80}N_6O_{18}Na$, 1411.5427; found, 1411.5497.

5.3.4.14. *6G-SMoC-Boc* **50**

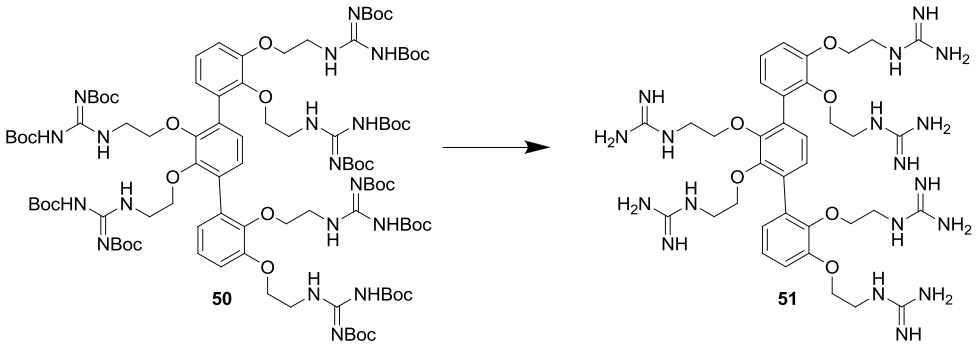


6-SMoC-NHCbz **48** (0.15 g, 0.11 mmol) was dissolved in DCM (4 mL) and HBr (30% in acetic acid, 1 mL) was added dropwise. After stirring at room temperature for 1.5 h, water (25 mL) was added to the mixture, the layers were separated and the aqueous layer was washed with DCM (2 x 25 mL). The water was then removed under vacuum and the crude hexa-amine was carefully dried under vacuum for several hours. The resulting solid was suspended in DCM (5 mL) and DIEA (1 mL) and *N*,*N*-di-boc-*N*'-trifluoromethanesulfonyl-guanidine (0.25 g, 0.63 mmol) were added. The mixture was stirred overnight at room temperature, diluted with DCM (20 mL) then washed with 2M $NaHSO_4$ (25 mL) followed by sat. aq. $NaHCO_3$ (25 mL) and brine (25 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated under vacuum. The residue was purified by flash chromatography using cyclohexane/EtOAc (85:15 to 65:35) as eluent to afford the title compound (0.12 g, 54% yield) as a white solid.

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.32-1.64 (m, 108H, 36 x Boc-CH_3), 3.42-3.78 (m, 8H, 4 x CH_2), 3.77-4.14 (m, 12H, 6 x CH_2), 4.11-4.31 (m, 4H, 2 x CH_2), 6.81-7.21 (m, 8H, 8 x ArH).

LRMS-ES (*m/z*): $(M + 2H)^{2+}$ calcd for $C_{96}H_{154}N_{18}O_{30}$, 1020.17; found, 1020.04.

5.3.4.15. *6G-SMoC* **51**



50 (20 mg, 0.01 mmol) was dissolved in a mixture of TFA/m-cresol (95:5, 5 ml) and stirred at room temperature for 2 h. The solvent was removed under vacuum and the product washed with ether (15 mg, 100% yield).

^{1}H-NMR (500 MHz, D_2O) δ: 2.99–3.17 (m, 8H), 3.46–3.60 (m, 4H), 3.75-3.90 (m, 8H), 4.05–4.23 (m, 4H), 6.77–6.90 (m, 2H), 6.95–7.22 (m, 6H).

HRMS-ES (*m/z*): $(M + H)^+$ calcd for $C_{36}H_{57}N_{18}O_6$, 837.4708; found, 837.4669.

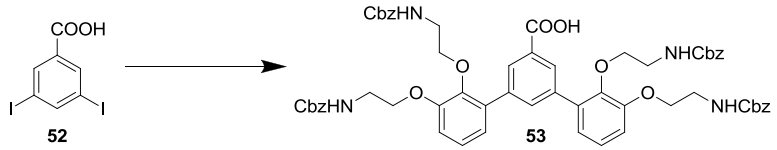
5.4. Chapter 4 Methods

5.4.1. Molecular dynamics

SMoC structures were imported into the YASARA[161] software as mol2 files. The starting structures were contained in a simulation cell which extended 5 Å larger than the SMoC structure along each axis *in vacuo*, the default pH set to 7.0 and the structures were minimised under the AMBER99[162] force field with default settings starting at 298 K for 1 ps.

5.4.2. Syntheses

5.4.2.1. *3,5-bis(2,3-bis(2-(((benzyloxy)carbonyl)amino)ethoxy)phenyl)benzoic acid* **53**



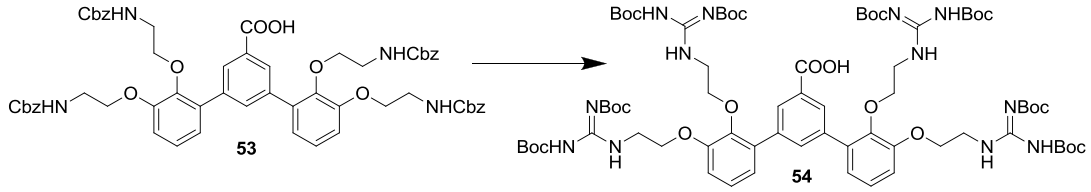
To a degassed solution of potassium (2,3-bis(2-(((benzyloxy)carbonyl)amino)-ethoxy)phenyl)trifluoroborate **18** (3.62 g, 6.34 mmol) in isopropanol/water (2/1, 21 ml) was added $PdCl_2dppf.CH_2Cl_2$ (123 mg, 0.15 mmol) and triethylamine (1.77 ml, 12.68 mmol) and the solution stirred for 2 minutes. 3,5-Diiodobenzoic acid **52** (789.1 mg, 2.11 mmol) was added and the reaction heated in a microwave reactor for 10 minutes at 75 °C. The reaction mixture was cooled, filtered and 1M HCl added (50 ml). The product was extracted with dichloromethane (3 x 25 ml) and the combined organic extracts were washed with brine (50 ml), dried over $MgSO_4$, filtered, and concentrated under vacuum. The mixture was purified by flash column chromatography using a gradient of 0-50% ethyl acetate in cyclohexane and the title product (1.7 g, 80% yield) obtained as a colourless oil.

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 3.27 (s, 4H, 2 x CH_2), 3.48-3.74 (m, 8H, 4 x CH_2), 4.11 (s, 4H, 2 x CH_2), 4.97-5.10 (m, 8H, 4 x Ph-O*CH*$_2$NH), 6.68-7.11 (m, 6H, 6 x ArH), 7.23-7.32 (m, 20H, 20 x Cbz-ArH), 7.66-8.29 (m, 3H, 3 x ArH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 40.81 (CH_2), 41.13 (CH_2), 66.87 (Cbz CH_2), 68.22 (CH_2), 113.43, 123.04, 124.90, 128.20 (Cbz ArC), 128.51 (Cbz ArC), 135.03, 136.44, 138.37, 145.06, 152.02, 156.77 (C=O).

HRMS-ES (*m/z*): $[M - H]^-$ calcd for $C_{59}H_{57}N_4O_{14}$, 1045.3871; found, 1045.3804.

5.4.2.2. *Tetra-Boc-4G-SMoC-dendrimer* **54**



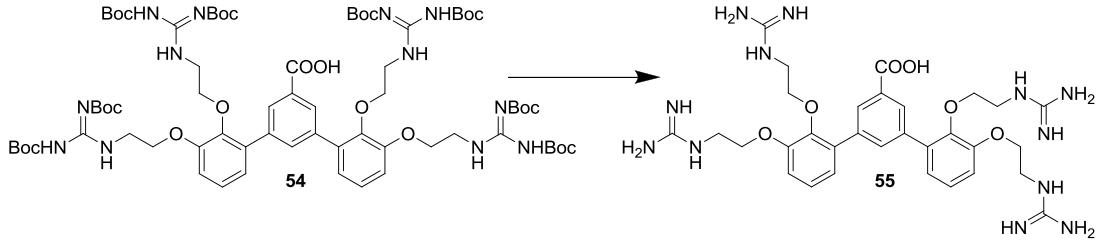
To a solution of 3,5-bis(2,3-bis(2-(((benzyloxy)carbonyl)amino)ethoxy)phenyl)-benzoic acid **53** (813 mg, 0.7 mmol) in dichloromethane (20 ml) was added 30% hydrogen bromide in acetic acid (5 ml) dropwise and the mixture stirred at room temperature for 3 h. The reaction was diluted with dichloromethane (20 ml) and the product extracted with water (3 x 20 ml). The aqueous extracts were carefully dried under vacuum for several hours. The crude amine was then dissolved in methanol (20 ml) and *N*,*N*′-Di-Boc-1*H*-pyrazole-1-carboxamidine (922 mg, 3.15 mmol) and DIEA (1.22 ml, 7 mmol) were added. The reaction was stirred at room temperature for 24 h under nitrogen. The solvent was removed under reduced pressure and the residue redissolved in ethyl acetate (50 ml) and 1M HCl added (50 ml). The layers were separated and the aqueous layer washed with ethyl acetate (3 x 20 ml). The combined organic extracts were washed with brine (50 ml), dried over $MgSO_4$, filtered, and concentrated under vacuum. The product was purified by flash column chromatography using a 0-25% ethyl acetate gradient in cyclohexane. The product was obtained as an off-white solid (651 mg, 63% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.42-1.51 (m, 72H, 24 x Boc-CH_3), 3.55 (m, 4H, 2 x CH_2), 3.87 (m, 4H, 2 x CH_2), 3.93 (m, 4H, 2 x CH_2), 4.23 (m, 4H, 2 x CH_2), 6.99 (m, 2H, 2 x ArH), 7.03 (m, 2H, 2 x ArH), 7.13 (m, 2H, 2 x ArH), 7.98 (m, 1H, ArH), 8.28 (m, 2H, 2 x ArH), 8.66 (m, 2H, 2 x NH), 8.81 (m, 2H, 2 x NH), 11.45 (bs, 1H, COOH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 28.06 (Boc-CH_3), 28.17 (Boc-CH_3), 28.36 (Boc-CH_3), 28.39 (Boc-CH_3), 53.49 (CH_2), 67.43 (CH_2), 79.06 (Boc *C*CH_3), 79.35 (Boc *C*CH_3), 82.80 (Boc *C*CH_3), 83.20 (Boc *C*CH_3), 113.31, 123.34, 124.77, 129.78, 135.12, 138.33, 145.16, 152.11, 153.01, 156.43, 163.54, 169.33 (COOH).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{71}H_{107}N_{12}O_{22}$, 1479.7423; found, 1479.7423.

5.4.2.3. *4G-SMoC-dendrimer* **55**

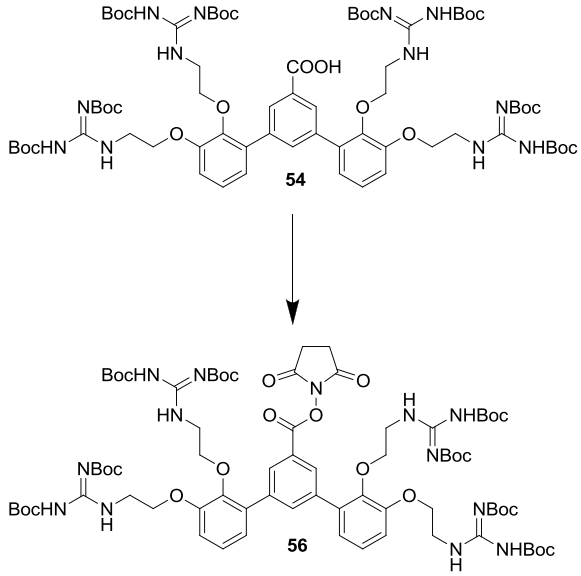


54 (30 mg, 0.02 mmol) was dissolved in DCM (2 mL) and TFA/*m*-cresol (95/5, 2 mL) was added and the mixture stirred for 4h at room temperature. The solvent was removed under vacuum and the residue washed three times with ether. The product was dissolved in water and lyophilised to form the tetra-TFA salt as a white solid (23 mg, 0.02 mmol, 100% yield).

^{1}H-NMR (500 MHz, D_2O) δ: 3.13 (m, 4H, 2 x CH_2), 3.65 (t, *J* = 5.0 Hz, 4H, 2 x CH_2), 3.92 (m, 4H, 2 x CH_2), 4.25 (t, *J* = 5.0 Hz, 4H, 2 x CH_2), 7.05 (dd, *J* = 1.4, 7.8 Hz, 2H, 2 x ArH), 7.12 (dd, *J* = 1.4, 8.3 Hz, 2H, 2 x ArH), 7.24 (t, *J* = 8.0 Hz, 2H, 2 x ArH), 7.80 (s, 1H, ArH), 8.11 (s, 2H, 2 x ArH).

HRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{31}H_{43}N_{12}O_6$, 679.3429; found, 679.3428.

5.4.2.4. *Tetra-Boc-4G-SMoC-dendrimer-succinimidyl ester* **56**



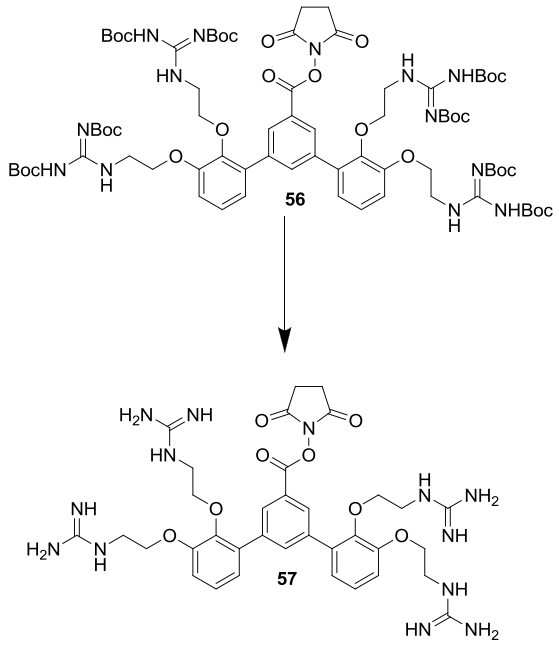
54 (100 mg, 0.07 mmol) was dissolved in DMF (10 ml) and 1-ethyl-3-(3-dimethyl-aminopropyl)carbodiimide hydrochloride (14 mg, 0.077 mmol) and *N*-hydroxysuccinimide (8.5 mg, 0.077 mmol) were added. The mixture was stirred for 17 h at room temperature under a nitrogen atmosphere. The solvent was removed under reduced pressure and the residue redissolved in ethyl acetate (20 ml), washed with water (3x10 ml), dried over $MgSO_4$, filtered and concentrated under vacuum. The product was purified by flash column chromatography over a 0-50% gradient of ethyl acetate in cyclohexane and obtained as a colourless oil (70 mg, 63% yield).

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 1.41-1.44 (m, 72H, 24Boc-CH_3), 2.89 (s, 4H, 2Su-CH_2), 3.51-3.61 (m, 4.9 Hz, 4H, 2CH_2), 4.00 – 3.80 (m, 8H, 4CH_2), 4.23 (t, *J* = 5.0 Hz, 4H, 2CH_2), 7.20 – 6.95 (m, 6H, 6Ar-CH), 8.11 (t, *J* = 1.6 Hz, 1H, Ar-H), 8.29 (d, *J* = 1.6 Hz, 2H, 2Ar-H), 11.44 (s, 2H, 2NH), 11.49 (s, 2H, 2NH).

^{13}C-NMR (500 MHz, $CDCl_3$) δ: 25.75 (Su-CH_2), 26.97 (Boc-CH_3), 28.17 (Boc-CH_3), 28.37 (Boc-CH_3), 40.06 (CH_2), 41.42 (CH_2), 67.47 (CH_2), 71.36 (CH_2), 78.90 (Boc-C), 79.29 (Boc-C), 82.70 (Boc-C), 83.17 (Boc-C), 113.64 (Ar-CH), 123.22 (Ar-CH), 124.88 (Ar-CH), 130.00 (Ar-CH), 134.49 (Ar-C), 138.91 (Ar-C), 145.13 (Ar-C), 152.16 (Boc C=O), 152.99 (Boc C=O), 156.19 (Boc C=O), 156.43 (Boc C=O), 161.88 (Guanidine C=N), 163.56 (Ester C=O), 169.09 (Su-C=O).

LRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{75}H_{110}N_{13}O_{24}$, 1576.8; found, 1576.8.

5.4.2.5. *4G-SMoC-dendrimer-succinimidyl ester* **57**

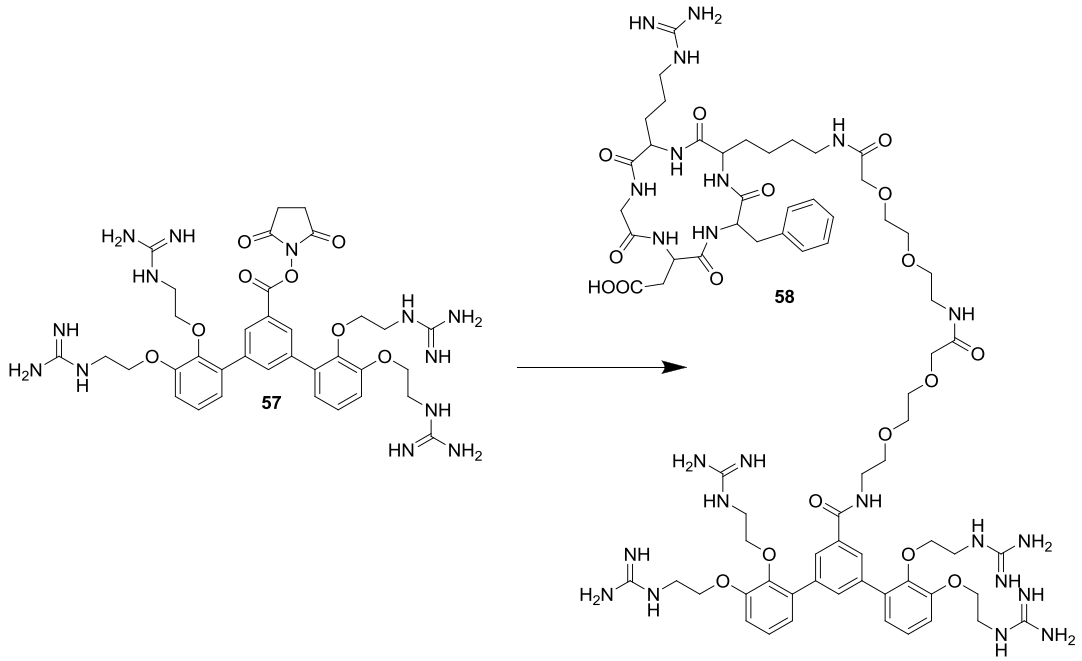


56 (96 mg, 0.061 mmol) was dissolved in DCM (2 mL) and TFA/*m*-cresol (95/5, 2 mL) was added, and stirred for 3h at room temperature. The solvent was removed under vacuum and the residue washed three times with ether. The product was dissolved in water and lyophilised to yield the tetra-TFA salt as a white solid (42 mg, 0.034 mmol, 56% yield).

^{1}H-NMR (500 Mhz, D_2O) δ: 3.07 (s, 4H, 2 x Su-CH_2), 3.33 – 3.15 (m, 4H, 2 x CH_2), 3.74 (t, *J* = 4.9 Hz, 4H, 2 x CH_2), 3.95-4.12 (m, 4H, 2 x CH_2), 4.34 (t, *J* = 4.9 Hz, 4H, 2 x CH_2), 7.29 – 7.08 (m, 4H, 4 x ArH), 7.43 – 7.29 (m, 2H, 2 x ArH), 8.01 (t, *J* = 1.5 Hz, 1H, ArH), 8.35 (t, *J* = 10.1 Hz, 2H, 2 x ArH).

LRMS-ES (*m/z*): $[M + H]^+$ calcd for $C_{35}H_{46}N_{13}O_8$, 776.4; found, 776.3.

5.4.2.6. *4G-SMoC-dendrimer-[PEG$_2$]$_2$-cRGDfK* **58**



57 (3 mg, 0.0024 mmol) was dissolved in DMF (200 µl) and *cyclo*[Arg-Gly-Asp-D-Phe-Lys(PEG-PEG)] (Peptides International, USA) (1.5 mg, 0.0024 mmol) and DIEA (5 µL, 0.00024 mmol) were added. The mixture was stirred at room temperature and monitored by LC-MS until all the peptide had been consumed. The product was purified by prep LCMS using a mobile phase of H_2O/MeCN (95/5). The product was dissolved in water and lyophilised to yield a white solid (1.5 mg, 0.0007 mmol, 30% yield).

LRMS-ES (*m/z*): $[M+H]^+$ calcd for $C_{70}H_{104}N_{23}O_{18}^+$, 1554.8; found, 1554.7. $[M+2H]^{2+}$ calcd for $C_{70}H_{104}N_{23}O_{18}^{2+}$, 777.9; found, 778.0.

HRMS-ES (*m/z*): $[M+3H]^{3+}$ calcd for $C_{70}H_{106}N_{23}O_{18}$, 518.5397; found, 518.9233.

CHAPTER 6

Discussion

6. Discussion

The extensive literature studies on the delivery of macromolecules such as proteins and oligonucleotides to the intracellular environment demonstrate that this is a prominent research question, with implications in the future of therapeutics. The discovery of the RNAi mechanism has opened up the possibility of effective gene therapy driven by siRNA-based drugs which act specifically and catalytically. However the high surface charge and molecular weight of siRNA has so far prohibited its delivery to its intracellular target. Viral delivery of gene therapies was previously thought to be the answer to this challenge, however safety concerns have led many researchers to seek alternative solutions. The discovery of CPPs showed that macromolecules may be transported across cell membranes via an unknown non-viral mechanism that appears to exist in a wide variety of cell types. However, CPPs alone are insufficient for high efficiency, targeted delivery of biological drugs *in vivo* due to their susceptibility to degradation, low bioavailability and lack of cell type specificity. Therefore, much attention has been given to the development of nanoparticles, built up from layers of different materials including lipids, peptides and targeting ligands, acting in concert to navigate their nucleic acid cargoes to the cytoplasms of their target cells.[118] This approach has been successful, with the application of nanoparticles demonstrated *in vivo* as demonstrated with SNALPs,[87] and also in clinical trials.[163] However, safety concerns still remain over nanoparticles, with variation in the particle size and charge having large impacts on their toxicity, as well as activation of the immune system.[164] Therefore, the search for a universal, safe and cost-effective delivery system for siRNA and other macromolecular drugs continues.

The work presented in this thesis represents significant progress in the development of the SMoC class of molecular transporters, which were first presented as protein delivery agents in 2007.[122] Tbe SMoCs may have significant advantages over nanoparticle methods for the delivery of macromolecules, due to their low cost of production and the possibility of overcoming the safety problems of more complex vectors. The previously published synthetic route to the SMoCs has been

optimised,[124] allowing higher yielding production of large quantities of compound at lower cost. The global yield of 4G-SMoC using the new reagents (Scheme 9) compared to the previous method (Scheme 1) has been increased 5-fold, and the trifluoroborate salt has allowed long-term storage of a convenient intermediate. Significant progress has also been made towards a new SMoC with a linker at the *meta* position to the biphenyl bond, which has previously been shown to increase 3G-SMoC uptake.[124]

Previously, SMoCs had only been shown to internalise proteins and dyes into cells. Since RNAi technology is becoming increasingly prevalent both as a research tool, and also in clinical trials, it was decided to test the ability of SMoCs to transfect siRNA into cells. Non-covalent complexes containing siRNA have been shown to be successfully internalised using guanidine-containing CPPs,[103-105, 107-108, 111, 114-117] and are preferable over covalently linked siRNA conjugates since they may be formed from non-modified siRNA and are easy to form, requiring no further synthetic steps. Non-covelent complexes of CPPs and siRNA are thought to result in protection of the siRNA from serum nucleases, as well as shielding the negative charges on the siRNA, therefore decreasing renal clearance and allowing passage through the lipophilic cell membrane. Stable SMoC-siRNA electrostatic complexes with micromolar binding affinities have been formed, which show for the first time that SMoCs may function as non-covalent as well as covalent macromolecular carriers in the same way as CPPs such as TAT and penetratin. 4G-SMoC has a significantly higher binding affinity than tetraarginine, perhaps due to the contribution from π-cation interactions. It would be interesting to compare SMoCs to penetratin to determine whether the amphipathic nature of the molecules account for this increased nucleic acid binding.

Human cell lines were transfected with SMoC-siRNA complexes or Lipofectamine, and 4G-SMoC resulted in a mRNA knockdown comparable to Lipofectamine. This result shows that SMoCs are able to transport active siRNA into cells in order to initiate RNAi-induced gene knockdown comparable to a commonly used laboratory

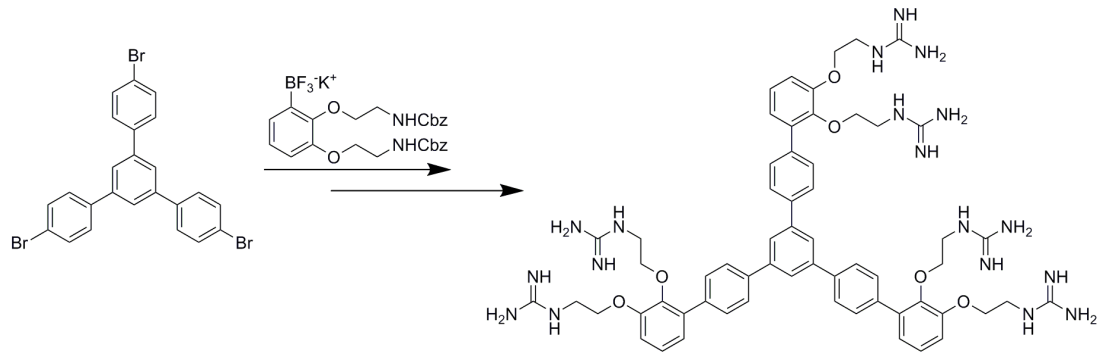
transfection reagent. This is particularly significant, as it expands the use of SMoCs beyond proteins, demonstrating that they may have potential as universal macromolecule delivery vectors for use in the laboratory and, with further development, biological therapeutics. The fact that no such commercial reagent which is able to transport both proteins and siRNA into cells currently exists makes these results particularly interesting, and it is proposed that SMoCs may form a new class of macromolecular transporter alongside viruses, CPPs and cationic liposomes.

In an attempt to enhance the uptake of SMoCs and increase the potency of the RNAi knockdown, a new SMoC was designed using computational techniques to maximise binding with cell surface GAGs, since it is proposed that GAG binding and clustering is a precursor to macropinocytosis.[38] This new compound represents the first in a new class of modular SMoCs which are designed to maximise anion complexation and synthesized using the new reagents and methods. It is also the first SMoC that deviates significantly from the standard biphenyl ring system that was modelled on a alpha-helical arginine-containing peptide. The high-yielding three-step synthetic route demonstrates how new compounds can easily be assembled using the new chemistry. In addition, the ability to form an activated ester of the final compound enables the new SMoC reagent to be stored and coupled to an appropriate cargo with ease at a later date. The mRNA knockdown produced using the new compound was not significantly larger than using the original 4G-SMoC, but was still comparable to Lipofectamine, confirming that these compounds may be used as an alternative to the currently available transfection reagents, particularly if they are able to overcome the difficulties of the current reagents, such as the high cytotoxicity of cationic liposome-based compounds.[80] The new compound has also been shown to be successfully coupled to a targeting peptide. The targeting of siRNA to specific cells is a key feature of successful delivery *in vivo*, where to achieve effective gene silencing at low dosage, specificity is vital. Targeting ligands are therefore considered to be an important component of siRNA delivery systems[118] and several such ligands have been identified for a variety of disease states, including tumour cells.[165] The new SMoC compound fulfills this criteria of siRNA delivery

systems, therby increasing the potential of SMoCs for use as macromolecular drug transporters.

6.1. Future Work

The work presented in this thesis has opened up further avenues of investigation for the development of SMoCs, both in the short term and long term. The synthesis of the dendrimer compound has demonstrated that modular SMoC compounds can easily be synthesized by coupling the trifluoroborate salt to aryl halides. This technique can be used to produce even larger macromolecular SMoC compounds, which may be used to encapsulate hydrophilic macromolecules for cell internalisation, in the same way that nanoparticles are formed using polymeric compounds such as cyclodextrin and PEI for siRNA delivery. For example, using the cheap, commercially available 1,3,5-tris(bromophenyl)benzene starting material, a large dendrimeric SMoC compound displaying six guanidine groups may be synthesized (Scheme 23). Such a molecule would have an increased number of positive charges as well as increased hydrophobicity, which may result in a stronger π-cation stabilising effect. In addition, by increasing the size of the SMoC molecule and the number charges, the compound may have a higher binding affinity for siRNA and cell surface GAGs, as well as being able to bind to the cell surface over a greater area, therefore possibly increasing uptake due to the lower entropic barrier. Analysis of the binding affinity of various SMoC compounds to heparin, including the new 4G-SMoC-Dr, is currently in progress using a surface plasmon resonance assay. The results from these experiments will indicate if the increased distance between guanidine groups is related to its affinity for anionic GAGs, as hypothesised.



Scheme 23. Proposed synthesis of a larger SMoC dendrimer from 1,3,5-tris(bromophenyl)-benzene using the potassium trifluoroborate intermediate. Subsequent guanidinylation with the new pyrazole reagent and deprotection would yield the product.

In the ABCD system of classifying siRNA nanocarriers described by Kostarelos and Miller,[118] the D component is a targeting molecule which enables the siRNA to be directed at a particular cell type by binding to specific cell surface receptors. As well as increasing the specificity of the siRNA delivery agents *in vivo*, targeting components may also aid uptake by increasing the local concentration at the cell membrane by binding to its target receptor, and therefore also lead to increased GAG and phospholipid binding which promotes uptake into cells. It was demonstrated in this thesis that the conjugation of an RGD peptide linked to a PEG linker is possible using the succinimidyl ester of the SMoC dendrimer. Using this activated ester, further conjugations to other targeting peptides may also be possible, and therefore SMoCs with specificity for different cell types may be produced. The conjugation of targeting peptides to SMoCs would be an interesting new approach, in order to introduce specificity as well as potential increased affinity for the cell membrane, and therefore perhaps greater potency.

One of the long term goals of SMoCs is for use in the delivery of macromolecular therapeutics. Many of the current clinical trials of RNAi use naked siRNA which is unable to be delivered systemically due to degradation by endogenous nucleases and rapid clearance due to its high hydrophilic nature. The use of cationic liposome-based laboratory transfection reagents is inappropriate for therapeutic use, due to

their high toxicity, and so there no widely used standard delivery vector for use *in vivo*. The development of a small molecule delivery vector based on SMoCs for siRNA and other macromolecules would be a significant achievement, and would offer many advantages over other compounds such as CPPs. For example, the ease and low cost of production resulting from the new SMoC synthetic modifications, as well as their ability to be coupled to a wide range of macromolecules either via the original disulfide linker, or peptide coupling via the new activated ester. The use of targeting peptides with SMoCs will allow for increased potency *in vivo* and the targeting of specific cell types and disease states, such as tumour cells overexpressing certain integrins. Therefore, *in vivo* testing of SMoCs is an important aim for the future of this work, and ultimately may allow SMoCs to be used for a wide range of applications.

Bibliography

Bibliography

1. Leader, B., Baca, Q. J., Golan, D. E. Protein therapeutics: A summary and pharmacological classification. *Nat. Rev. Drug Discov.* **2008,** *7* (1), 21-39.

2. Leaman, D. W. Recent progress in oligonucleotide therapeutics: antisense to aptamers. *Expert Opin. Drug Disc.* **2008,** *3* (9), 997-1009.

3. Carter, P. J. Introduction to current and future protein therapeutics: A protein engineering perspective. *Exp. Cell Res.* **2011,** *317* (9), 1261-1269.

4. Waldmann, T. A. Immunotherapy: past, present and future. *Nat. Med.* **2003,** *9* (3), 269-277.

5. Reichert, J. M. Antibody-based therapeutics to watch in 2011. *mAbs* **2011,** *3* (1), 76-99.

6. Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., Mello, C. C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **1998,** *391* (6669), 806-811.

7. de Fougerolles, A., Vornlocher, H. P., Maraganore, J., Lieberman, J. Interfering with disease: a progress report on siRNA-based therapeutics. *Nat. Rev. Drug Discov.* **2007,** *6* (6), 443-453.

8. Guo, J., Fisher, K. A., Darcy, R., Cryan, J. F., O'Driscoll, C. Therapeutic targeting in the silent era: advances in non-viral siRNA delivery. *Mol. Biosyst.* **2010,** *6* (7), 1143-61.

9. Giezen, T. J., Mantel-Teeuwisse, A. K., Straus, S. M. J. M., Schellekens, H., Leufkens, H. G. M., Egberts, A. C. G. Safety-Related Regulatory Actions for Biologicals Approved in the United States and the European Union. *J. Am. Med. Assoc.* **2008,** *300* (16), 1887-1896.

10. Juliano, R. Challenges to macromolecular drug delivery. *Biochem. Soc. Trans.* **2007,** *35*, 41-43.

11. Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **2001,** *46* (1-3), 3-26.

12. Belting, M., Wittrup, A. Macromolecular drug delivery: basic principles and therapeutic applications. *Mol. Biotechnol.* **2009,** *43* (1), 89-94.

13. Frankel, A. D., Pabo, C. O. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* **1988,** *55* (6), 1189-93.

14. Joliot, A., Pernelle, C., Deagostini-Bazin, H., Prochiantz, A. Antennapedia homeobox peptide regulates neural morphogenesis. *Proc. Natl. Acad. Sci. USA* **1991,** *88* (5), 1864-8.

15. Derossi, D., Joliot, A. H., Chassaing, G., Prochiantz, A. The third helix of the Antennapedia homeodomain translocates through biological membranes. *J. Biol. Chem.* **1994,** *269* (14), 10444-10450.

16. Vivès, E., Brodin, P., Lebleu, B. A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J. Biol. Chem.* **1997,** *272* (25), 16010-7.

17. Morris, M. C., Vidal, P., Chaloin, L., Heitz, F., Divita, G. A new peptide vector for efficient delivery of oligonucleotides into mammalian cells. *Nucleic Acids Res.* **1997,** *25* (14), 2730-2736.

18. Schwarze, S. R., Ho, A., Vocero-Akbani, A., Dowdy, S. F. *In vivo* Protein Transduction: Delivery of a Biologically Active Protein into the Mouse. *Science* **1999,** *285* (5433), 1569-1572.

19. Morris, M. C., Depollier, J., Mery, J., Heitz, F., Divita, G. A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nat. Biotechnol.* **2001,** *19* (12), 1173-1176.

20. Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L. L., Pepinsky, B., Barsoum, J. Tat-Mediated Delivery of Heterologous Proteins into Cells. *Proc. Natl. Acad. Sci. USA* **1994,** *91* (2), 664-668.

21. Ruan, G., Agrawal, A., Marcus, A. I., Nie, S. Imaging and tracking of tat peptide-conjugated quantum dots in living cells: new insights into nanoparticle uptake, intracellular transport, and vesicle shedding. *J. Am. Chem. Soc.* **2007,** *129* (47), 14759-14766.

22. Abes, R., Arzumanov, A. A., Moulton, H. M., Abes, S., Lvanciva, G. D., Lversen, P. L., Gait, M. J., Lebleu, B. Cell-penetrating-peptide-based delivery of oligonucleotides: an overview. *Biochem. Soc. Trans.* **2007,** *35*, 775-779.

23. Juliano, R., Alam, M. R., Dixit, V., Kang, H. Mechanisms and strategies for effective delivery of antisense and siRNA oligonucleotides. *Nucleic Acids Res.* **2008,** *36* (12), 4158-4171.

24. Derossi, D., Calvet, S., Trembleau, A., Brunissen, A., Chassaing, G., Prochiantz, A. Cell Internalization of the Third Helix of the Antennapedia Homeodomain Is Receptor-independent. *J. Biol. Chem.* **1996,** *271* (30), 18188-18193.

25. Richard, J. P., Melikov, K., Vives, E., Ramos, C., Verbeure, B., Gait, M. J., Chernomordik, L. V., Lebleu, B. Cell-penetrating peptides - A reevaluation of the mechanism of cellular uptake. *J. Biol. Chem.* **2003,** *278* (1), 585-590.

26. Lundberg, M., Wikstrom, S., Johansson, M. Cell Surface Adherence and Endocytosis of Protein Transduction Domains. *Mol. Ther.* **2003,** *8* (1), 143-150.

27. Wadia, J. S., Stan, R. V., Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **2004,** *10* (3), 310-315.

28. Nakase, I., Niwa, M., Takeuchi, T., Sonomura, K., Kawabata, N., Koike, Y., Takehashi, M., Tanaka, S., Ueda, K., Simpson, J. C., Jones, A. T., Sugiura, Y., Futaki, S. Cellular Uptake of Arginine-Rich Peptides: Roles for Macropinocytosis and Actin Rearrangement. *Mol. Ther.* **2004,** *10* (6), 1011-1022.

29. Jiao, C. Y., Delaroche, D., Burlina, F., Alves, I. D., Chassaing, G., Sagan, S. Translocation and Endocytosis for Cell-penetrating Peptide Internalization. *J. Biol. Chem.* **2009,** *284* (49), 33957-33965.

30. Guterstam, P., Madani, F., Hirose, H., Takeuchi, T., Futaki, S., El Andaloussi, S., Gräslund, A., Langel, U. Elucidating cell-penetrating peptide mechanisms of action for membrane interaction, cellular uptake, and translocation utilizing the hydrophobic counter-anion pyrenebutyrate. *Biochim. Biophys. Acta* **2009,** *1788* (12), 2509-17.

31. Nakase, I., Tadokoro, A., Kawabata, N., Takeuchi, T., Katoh, H., Hiramoto, K., Negishi, M., Nomizu, M., Sugiura, Y., Futaki, S. Interaction of Arginine-Rich Peptides with Membrane-Associated Proteoglycans Is Crucial for Induction of Actin Organization and Macropinocytosis†. *Biochemistry* **2006,** *46* (2), 492-501.

32. Simon, M. J., Gao, S., Kang, W. H., Banta, S., Morrison, B. TAT-mediated intracellular protein delivery to primary brain cells is dependent on glycosaminoglycan expression. *Biotechnol. Bioeng.* **2009,** *104* (1), 10-9.

33. Patel, L. N., Zaro, J. L., Shen, W. C. Cell penetrating peptides: Intracellular pathways and pharmaceutical perspectives. *Pharm. Res.* **2007,** *24* (11), 1977-1992.

34. Fuchs, S. M., Raines, R. T. Pathway for Polyarginine Entry into Mammalian Cells†. *Biochemistry* **2004,** *43* (9), 2438-2444.

35. Gerbal-chaloin, S., Gondeau, C., Aldrian-herrada, G., Heitz, F., Gauthier-rouvière, C., Divita, G. First step of the cell-penetrating peptide mechanism involves Rac1 GTPase-dependent actin-network remodelling. *Biol. Cell* **2007,** *099* (4), 223-238.

36. Couchman, J. R. Syndecans: proteoglycan regulators of cell-surface microdomains? *Nat. Rev. Mol. Cell Biol.* **2003,** *4* (12), 926-938.

37. Conner, S. D., Schmid, S. L. Regulated portals of entry into the cell. *Nature* **2003,** *422* (6927), 37-44.

38. Ziegler, A. Thermodynamic studies and binding mechanisms of cell-penetrating peptides with lipids and glycosaminoglycans. *Adv. Drug Del. Rev.* **2008,** *60* (4-5), 580-597.

39. Ziegler, A., Seelig, J. Interaction of the Protein Transduction Domain of HIV-1 TAT with Heparan Sulfate: Binding Mechanism and Thermodynamic Parameters. *Biophys. J.* **2004,** *86* (1), 254-263.

40. Ziegler, A., Seelig, J. Binding and Clustering of Glycosaminoglycans: A Common Property of Mono- and Multivalent Cell-Penetrating Compounds. *Biophys. J.* **2008,** *94* (6), 2142-2149.

41. Gump, J. M., June, R. K., Dowdy, S. F. Revised role of glycosaminoglycans in TAT protein transduction domain-mediated cellular transduction. *J. Biol. Chem.* **2010,** *285* (2), 1500-7.

42. Esbjorner, E. K., Lincoln, P., Norden, B. Counterion-mediated membrane penetration: Cationic cell-penetrating peptides overcome Born energy barrier by ion-pairing with phospholipids. *BBA-Biomembranes* **2007,** *1768* (6), 1550-1558.

43. Rothbard, J. B., Jessop, T. C., Wender, P. A. Adaptive translocation: the role of hydrogen bonding and membrane potential in the uptake of guanidinium-rich transporters into cells. *Adv. Drug Del. Rev.* **2005,** *57* (4), 495-504.

44. El-Sayed, A., Khalil, I. A., Kogure, K., Futaki, S., Harashima, H. Octaarginine- and Octalysine-modified Nanoparticles Have Different Modes of Endosomal Escape. *J. Biol. Chem.* **2008,** *283* (34), 23450-23461.

45. Sakai, N., Matile, S. Anion-Mediated Transfer of Polyarginine across Liquid and Bilayer Membranes. *J. Am. Chem. Soc.* **2003,** *125* (47), 14348-14356.

46. Mishra, A., Gordon, V. D., Yang, L., Coridan, R., Wong, G. C. HIV TAT forms pores in membranes by inducing saddle-splay curvature: potential role of bidentate hydrogen bonding. *Angew. Chem. Int. Ed.* **2008,** *47* (16), 2986-9.

47. Lamazière, A., Wolf, C., Lambert, O., Chassaing, G., Trugnan, G., Ayala-Sanmartin, J. The homeodomain derived peptide Penetratin induces curvature of fluid membrane domains. *PLoS One* **2008,** *3* (4), e1938.

48. Yezid, H., Konate, K., Debaisieux, S., Bonhoure, A., Beaumelle, B. Mechanism for HIV-1 Tat insertion into the endosome membrane. *J. Biol. Chem.* **2009,** *284* (34), 22736-46.

49. Sontheimer, E. J. Assembly and function of RNA silencing complexes. *Nat. Rev. Mol. Cell Biol.* **2005,** *6* (2), 127-138.

50. Goyal, B. R., Patel, M. M., Soni, M. K., Bhadada, S. V. Therapeutic opportunities of small interfering RNA. *Fundam. Clin. Pharmacol.* **2009,** *23* (4), 367-86.

51. Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **2001,** *411* (6836), 494-498.

52. Hamilton, A. J., Baulcombe, D. C. A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science* **1999,** *286* (5441), 950-952.

53. Bernstein, E., Caudy, A. A., Hammond, S. M., Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **2001,** *409* (6818), 363-366.

54. Hammond, S. M., Bernstein, E., Beach, D., Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **2000,** *404* (6775), 293-296.

55. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., Hannon, G. J. Argonaute2, a Link Between Genetic and Biochemical Analyses of RNAi. *Science* **2001,** *293* (5532), 1146-1150.

56. Collins, R. E., Cheng, X. D. Structural domains in RNAi. *FEBS Lett.* **2005,** *579* (26), 5841-5849.

57. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P., Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **2005,** *123* (4), 607-620.

58. Liu, Q., Rand, T. A., Kalidas, S., Du, F., Kim, H.-E., Smith, D. P., Wang, X. R2D2, a Bridge Between the Initiation and Effector Steps of the Drosophila RNAi Pathway. *Science* **2003,** *301* (5641), 1921-1925.

59. Paroo, Z., Liu, Q. H., Wang, X. D. Biochemical mechanisms of the RNA-induced silencing complex. *Cell Res.* **2007,** *17* (3), 187-194.

60. Tomari, Y., Matranga, C., Haley, B., Martinez, N., Zamore, P. D. A Protein Sensor for siRNA Asymmetry. *Science* **2004,** *306* (5700), 1377-1380.

61. Rand, T. A., Petersen, S., Du, F., Wang, X. Argonaute2 Cleaves the Anti-Guide Strand of siRNA during RISC Activation. *Cell* **2005,** *123* (4), 621-629.

62. Song, J.-J., Smith, S. K., Hannon, G. J., Joshua-Tor, L. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science* **2004,** *305* (5689), 1434-1437.

63. Rivas, F. V., Tolia, N. H., Song, J.-J., Aragon, J. P., Liu, J., Hannon, G. J., Joshua-Tor, L. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.* **2005,** *12* (4), 340-349.

64. Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., Liu, Q. Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Genes Dev.* **2005,** *19* (14), 1674-1679.

65. Okamura, K., Ishizuka, A., Siomi, H., Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **2004,** *18* (14), 1655-1666.

66. Meister, G., Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **2004,** *431* (7006), 343-349.

67. Shim, M. S., Kwon, Y. J. Efficient and targeted delivery of siRNA *in vivo*. *FEBS J.* **2010,** *277* (23), 4814-27.

68. Whitehead, K., Langer, R., Anderson, D. Knocking down barriers: advances in siRNA delivery. *Nat. Rev. Drug Discov.* **2009,** *8* (2), 129-138.

69. Barquinero, J., Eixarch, H., Perez-Melgosa, M. Retroviral vectors: new applications for an old tool. *Gene Ther.* **2004,** *11* (S1), S3-S9.

70. Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., Elbashir, S., Geick, A., Hadwiger, P., Harborth, J., John, M., Kesavan, V., Lavine, G., Pandey, R. K., Racie, T., Rajeev, K. G., Rohl, I., Toudjarska, I., Wang, G., Wuschko, S., Bumcrot, D., Koteliansky, V., Limmer, S., Manoharan, M., Vornlocher, H.-P. Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **2004,** *432* (7014), 173-178.

71. Wolfrum, C., Shi, S., Jayaprakash, K. N., Jayaraman, M., Wang, G., Pandey, R. K., Rajeev, K. G., Nakayama, T., Charrise, K., Ndungo, E. M., Zimmermann, T., Koteliansky, V., Manoharan, M., Stoffel, M. Mechanisms and optimization of *in vivo* delivery of lipophilic siRNAs. *Nat. Biotechnol.* **2007,** *25* (10), 1149-1157.

72. Dassie, J. P., Liu, X.-y., Thomas, G. S., Whitaker, R. M., Thiel, K. W., Stockdale, K. R., Meyerholz, D. K., McCaffrey, A. P., McNamara, J. O., Giangrande, P. H. Systemic administration of optimized aptamer-siRNA chimeras promotes regression of PSMA-expressing tumors. *Nat. Biotechnol.* **2009,** *27* (9), 839-846.

73. Felgner, P. L., Gadek, T. R., Holm, M., Roman, R., Chan, H. W., Wenz, M., Northrop, J. P., Ringold, G. M., Danielsen, M. Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci. USA* **1987,** *84* (21), 7413-7.

74. Malone, R. W., Felgner, P. L., Verma, I. M. Cationic liposome-mediated RNA transfection. *Proc. Natl. Acad. Sci. USA* **1989,** *86* (16), 6077-81.

75. Behr, J. P., Demeneix, B., Loeffler, J. P., Perez-Mutul, J. Efficient gene transfer into mammalian primary endocrine cells with lipopolyamine-coated DNA. *Proc. Natl. Acad. Sci. USA* **1989,** *86* (18), 6982-6.

76. Gao, X., Huang, L. A novel cationic liposome reagent for efficient transfection of mammalian cells. *Biochem. Biophys. Res. Commun.* **1991,** *179* (1), 280-5.

77. Rose, J. K., Buonocore, L., Whitt, M. A. A new cationic liposome reagent mediating nearly quantitative transfection of animal cells. *BioTechniques* **1991,** *10* (4), 520-5.

78. San, H., Yang, Z. Y., Pompili, V. J., Jaffe, M. L., Plautz, G. E., Xu, L., Felgner, J. H., Wheeler, C. J., Felgner, P. L., Gao, X. Safety and short-term toxicity of a novel cationic lipid formulation for human gene therapy. *Hum. Gene Ther.* **1993,** *4* (6), 781-8.

79. Omidi, Y., Hollins, A. J., Benboubetra, M., Drayton, R., Benter, I. F., Akhtar, S. Toxicogenomics of Non-viral Vectors for Gene Therapy: A Microarray Study of Lipofectin- and Oligofectamine-induced Gene Expression Changes in Human Epithelial Cells. *J. Drug Targeting* **2003,** *11* (6), 311-323.

80. Kedmi, R., Ben-Arie, N., Peer, D. The systemic toxicity of positively charged lipid nanoparticles and the role of Toll-like receptor 4 in immune activation. *Biomaterials* **2010,** *31* (26), 6867-6875.

81. Farhood, H., Serbina, N., Huang, L. The role of dioleoyl phosphatidylethanolamine in cationic liposome mediated gene transfer. *BBA-Biomembranes* **1995,** *1235* (2), 289-295.

82. Xu, L., Anchordoquy, T. J. Cholesterol domains in cationic lipid/DNA complexes improve transfection. *BBA-Biomembranes* **2008,** *1778* (10), 2177-2181.

83. Siegel, D. P., Epand, R. M. The mechanism of lamellar-to-inverted hexagonal phase transitions in phosphatidylethanolamine: implications for membrane fusion mechanisms. *Biophys. J.* **1997,** *73* (6), 3089-3111.

84. Simberg, D., Weisman, S., Talmon, Y., Faerman, A., Shoshani, T., Barenholz, Y. The Role of Organ Vascularization and Lipoplex-Serum Initial Contact in Intravenous Murine Lipofection. *J. Biol. Chem.* **2003,** *278* (41), 39858-39865.

85. Hirsch-Lerner, D., Zhang, M., Eliyahu, H., Ferrari, M. E., Wheeler, C. J., Barenholz, Y. Effect of "helper lipid" on lipoplex electrostatics. *BBA-Biomembranes* **2005,** *1714* (2), 71-84.

86. Morrissey, D. V., Lockridge, J. A., Shaw, L., Blanchard, K., Jensen, K., Breen, W., Hartsough, K., Machemer, L., Radka, S., Jadhav, V., Vaish, N., Zinnen, S., Vargeese, C., Bowman, K., Shaffer, C. S., Jeffs, L. B., Judge, A., MacLachlan, I., Polisky, B. Potent and persistent *in vivo* anti-HBV activity of chemically modified siRNAs. *Nat. Biotechnol.* **2005,** *23* (8), 1002-1007.

87. Zimmermann, T. S., Lee, A. C. H., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M. N., Harborth, J., Heyes, J. A., Jeffs, L. B., John, M., Judge, A. D., Lam, K., McClintock, K., Nechev, L. V., Palmer, L. R., Racie, T., Röhl, I., Seiffert, S., Shanmugam, S., Sood, V., Soutschek, J., Toudjarska, I., Wheat, A. J., Yaworski, E., Zedalis, W., Koteliansky, V., Manoharan, M., Vornlocher, H.-P., MacLachlan, I. RNAi-mediated gene silencing in non-human primates. *Nature* **2006,** *441* (7089), 111-114.

88. Geisbert, T. W., Lee, A. C. H., Robbins, M., Geisbert, J. B., Honko, A. N., Sood, V., Johnson, J. C., de Jong, S., Tavakoli, I., Judge, A., Hensley, L. E., MacLachlan, I. Postexposure protection of non-human primates against a lethal Ebola virus challenge with RNA interference: a proof-of-concept study. *Lancet* **2010,** *375* (9729), 1896-1905.

89. Semple, S. C., Akinc, A., Chen, J., Sandhu, A. P., Mui, B. L., Cho, C. K., Sah, D. W. Y., Stebbing, D., Crosley, E. J., Yaworski, E., Hafez, I. M., Dorkin, J. R., Qin, J., Lam, K., Rajeev, K. G., Wong, K. F., Jeffs, L. B., Nechev, L., Eisenhardt, M. L., Jayaraman, M., Kazem, M., Maier, M. A., Srinivasulu, M., Weinstein, M. J., Chen, Q., Alvarez, R., Barros, S. A., De, S., Klimuk, S. K., Borland, T., Kosovrasti, V., Cantley, W. L., Tam, Y. K., Manoharan, M., Ciufolini, M. A., Tracy, M. A., de Fougerolles, A., MacLachlan, I., Cullis, P. R., Madden, T. D., Hope, M. J. Rational design of cationic lipids for siRNA delivery. *Nat. Biotechnol.* **2010,** *28* (2), 172-176.

90. Mével, M., Kamaly, N., Carmona, S., Oliver, M. H., Jorgensen, M. R., Crowther, C., Salazar, F. H., Marion, P. L., Fujino, M., Natori, Y., Thanou, M., Arbuthnot, P., Yaouanc, J.-J., Jaffrès, P. A., Miller, A. D. DODAG; a versatile new cationic lipid that mediates efficient delivery of pDNA and siRNA. *J. Controlled Release* **2010,** *143* (2), 222-232.

91. Klein, E., Leborgne, C., Ciobanu, M., Klein, J., Frisch, B., Pons, F., Zuber, G., Scherman, D., Kichler, A., Lebeau, L. Nucleic acid transfer with hemifluorinated polycationic lipids. *Biomaterials* **2010,** *31* (17), 4781-4788.

92. Zhu, L., Mahato, R. I. Lipid and polymeric carrier-mediated nucleic acid delivery. *Expert Opin. Drug Del.* **2010,** *7* (10), 1209-26.

93. Urban-Klein, B., Werth, S., Abuharbeid, S., Czubayko, F., Aigner, A. RNAi-mediated gene-targeting through systemic application of polyethylenimine (PEI)-complexed siRNA *in vivo*. *Gene Ther.* **2004,** *12* (5), 461-466.

94. Sonawane, N. D., Szoka, F. C., Verkman, A. S. Chloride Accumulation and Swelling in Endosomes Enhances DNA Transfer by Polyamine-DNA Polyplexes. *J. Biol. Chem.* **2003,** *278* (45), 44826-44831.

95. Thomas, M., Lu, J. J., Ge, Q., Zhang, C., Chen, J., Klibanov, A. M. Full deacylation of polyethylenimine dramatically boosts its gene delivery efficiency and specificity to mouse lung. *Proc. Natl. Acad. Sci. USA* **2005,** *102* (16), 5679-5684.

96. Fischer, D., Li, Y., Ahlemeyer, B., Krieglstein, J., Kissel, T. *In vitro* cytotoxicity testing of polycations: influence of polymer structure on cell viability and hemolysis. *Biomaterials* **2003,** *24* (7), 1121-1131.

97. Li, J., Loh, X. J. Cyclodextrin-based supramolecular architectures: Syntheses, structures, and applications for drug and gene delivery. *Adv. Drug Del. Rev.* **2008,** *60* (9), 1000-1017.

98. Muratovska, A., Eccles, M. R. Conjugate for efficient delivery of short interfering RNA (siRNA) into mammalian cells. *FEBS Lett.* **2004,** *558* (1-3), 63-68.

99. Meade, B. R., Dowdy, S. F. Exogenous siRNA delivery using peptide transduction domains/cell penetrating peptides. *Adv. Drug Del. Rev.* **2007,** *59* (2-3), 134-140.

100. Chiu, Y.-L., Ali, A., Chu, C.-y., Cao, H., Rana, T. M. Visualizing a Correlation between siRNA Localization, Cellular Uptake, and RNAi in Living Cells. *Chem. Biol.* **2004,** *11* (8), 1165-1175.

101. Davidson, T. J., Harel, S., Arboleda, V. A., Prunell, G. F., Shelanski, M. L., Greene, L. A., Troy, C. M. Highly Efficient Small Interfering RNA Delivery to Primary Mammalian Neurons Induces MicroRNA-Like Effects before mRNA Degradation. *J. Neurosci.* **2004,** *24* (45), 10040-10046.

102. Moschos, S. A., Jones, S. W., Perry, M. M., Williams, A. E., Erjefalt, J. S., Turner, J. J., Barnes, P. J., Sproat, B. S., Gait, M. J., Lindsay, M. A. Lung Delivery Studies Using siRNA Conjugated to TAT(48−60) and Penetratin Reveal Peptide Induced Reduction in Gene Expression and Induction of Innate Immunity. *Bioconj. Chem.* **2007,** *18* (5), 1450-1459.

103. Simeoni, F., Morris, M. C., Heitz, F., Divita, G. Insight into the mechanism of the peptide-based gene delivery system MPG: implications for delivery of siRNA into mammalian cells. *Nucleic Acids Res.* **2003,** *31* (11), 2717-2724.

104. Veldhoen, S., Laufer, S. D., Trampe, A., Restle, T. Cellular delivery of small interfering RNA by a non-covalently attached cell-penetrating peptide: quantitative analysis of uptake and biological effect. *Nucleic Acids Res.* **2006,** *34* (22), 6561-6573.

105. Lundberg, P., El Andaloussi, S., Sutlu, T., Johansson, H., Langel, U. Delivery of short interfering RNA using endosomolytic cell-penetrating peptides. *FASEB J.* **2007,** *21* (11), 2664-2671.

106. Magzoub, M., Pramanik, A., Graslund, A. Modeling the endosomal escape of cell-penetrating peptides: Transmembrane pH gradient driven translocation across phospholipid bilayers. *Biochemistry* **2005,** *44* (45), 14890-14897.

107. El Andaloussi, S., Lehto, T., Mager, I., Rosenthal-Aizman, K., Oprea, II, Simonson, O. E., Sork, H., Ezzat, K., Copolovici, D. M., Kurrikoff, K., Viola, J. R.,

Zaghloul, E. M., Sillard, R., Johansson, H. J., Said Hassane, F., Guterstam, P., Suhorutsenko, J., Moreno, P. M., Oskolkov, N., Halldin, J., Tedebark, U., Metspalu, A., Lebleu, B., Lehtio, J., Smith, C. I., Langel, U. Design of a peptide-based vector, PepFect6, for efficient delivery of siRNA in cell culture and systemically *in vivo*. *Nucleic Acids Res.* **2011,** *39* (9), 3972-3987.

108. Crombez, L., Aldrian-Herrada, G., Konate, K., Nguyen, Q. N., McMaster, G. K., Brasseur, R., Heitz, F., Divita, G. A new potent secondary amphipathic cell-penetrating peptide for siRNA delivery into mammalian cells. *Mol. Ther.* **2009,** *17* (1), 95-103.

109. Konate, K., Crombez, L., Deshayes, S., Decaffmeyer, M., Thomas, A., Brasseur, R., Aldrian, G., Heitz, F., Divita, G. Insight into the cellular uptake mechanism of a secondary amphipathic cell-penetrating peptide for siRNA delivery. *Biochemistry* **2010,** *49* (16), 3393-402.

110. Unnamalai, N., Kang, B. G., Lee, W. S. Cationic oligopeptide-mediated delivery of dsRNA for post-transcriptional gene silencing in plant cells. *FEBS Lett.* **2004,** *566* (1-3), 307-310.

111. Kim, S. W., Kim, N. Y., Choi, Y. B., Park, S. H., Yang, J. M., Shin, S. RNA interference *in vitro* and *in vivo* using an arginine peptide/siRNA complex system. *J. Controlled Release* **2010,** *143* (3), 335-43.

112. Arthanari, Y., Pluen, A., Rajendran, R., Aojula, H., Demonacos, C. Delivery of therapeutic shRNA and siRNA by Tat fusion peptide targeting BCR-ABL fusion gene in Chronic Myeloid Leukemia cells. *J. Controlled Release* **2010,** *145* (3), 272-80.

113. Ishihara, T., Goto, M., Kodera, K., Kanazawa, H., Murakami, Y., Mizushima, Y., Higaki, M. Intracellular delivery of siRNA by cell-penetrating peptides modified with cationic oligopeptides. *Drug Deliv.* **2009,** *16* (3), 153-159.

114. Endoh, T., Sisido, M., Ohtsuki, T. Cellular siRNA Delivery Mediated by a Cell-Permeant RNA-Binding Protein and Photoinduced RNA Interference. *Bioconj. Chem.* **2008,** *19* (5), 1017-1024.

115. Eguchi, A., Meade, B. R., Chang, Y. C., Fredrickson, C. T., Willert, K., Puri, N., Dowdy, S. F. Efficient siRNA delivery into primary cells by a peptide transduction domain-dsRNA binding domain fusion protein. *Nat. Biotechnol.* **2009,** *27* (6), 567-U110.

116. Crombez, L., Morris, M. C., Dufort, S., Aldrian-Herrada, G., Nguyen, Q., Mc Master, G., Coll, J.-L., Heitz, F., Divita, G. Targeting cyclin B1 through peptide-based delivery of siRNA prevents tumour growth. *Nucleic Acids Res.* **2009,** *37* (14), 4559-4569.

117. Kim, W. J., Christensen, L. V., Jo, S., Yockman, J. W., Jeong, J. H., Kim, Y.-H., Kim, S. W. Cholesteryl Oligoarginine Delivering Vascular Endothelial Growth

Factor siRNA Effectively Inhibits Tumor Growth in Colon Adenocarcinoma. *Mol. Ther.* **2006,** *14* (3), 343-350.

118. Kostarelos, K., Miller, A. D. Synthetic, self-assembly ABCD nanoparticles; a structural paradigm for viable synthetic non-viral vectors. *Chem. Soc. Rev.* **2005,** *34* (11), 970-994.

119. Rao, N. M. Cationic lipid-mediated nucleic acid delivery: beyond being cationic. *Chem. Phys. Lipids* **2010,** *163* (3), 245-52.

120. Li, W., Szoka, F. C. Lipid-based nanoparticles for nucleic acid delivery. *Pharm. Res.* **2007,** *24* (3), 438-49.

121. Rozema, D. B., Lewis, D. L., Wakefield, D. H., Wong, S. C., Klein, J. J., Roesch, P. L., Bertin, S. L., Reppen, T. W., Chu, Q., Blokhin, A. V., Hagstrom, J. E., Wolff, J. A. Dynamic PolyConjugates for targeted *in vivo* delivery of siRNA to hepatocytes. *Proc. Natl. Acad. Sci. USA* **2007,** *104* (32), 12982-12987.

122. Okuyama, M., Laman, H., Kingsbury, S. R., Visintin, C., Leo, E., Eward, K. L., Stoeber, K., Boshoff, C., Williams, G. H., Selwood, D. L. Small-molecule mimics of an [alpha]-helix for efficient transport of proteins into cells. *Nat. Methods* **2007,** *4* (2), 153-159.

123. Jacoby, E. Biphenyls as potential mimetics of protein [alpha]-helix. *Bioorg. Med. Chem. Lett.* **2002,** *12* (6), 891-893.

124. Rebstock, A. S., Visintin, C., Leo, E., Posada, C. G., Kingsbury, S. R., Williams, G. H., Stoeber, K., Selwood, D. L. Modular assembly using sequential palladium coupling gives easy access to the SMoC class of cellular transporters. *ChemBioChem* **2008,** *9* (11), 1787-1796.

125. Wohlschlegel, J. A., Dwyer, B. T., Dhar, S. K., Cvetic, C., Walter, J. C., Dutta, A. Inhibition of Eukaryotic DNA Replication by Geminin Binding to Cdt1. *Science* **2000,** *290* (5500), 2309-2312.

126. Molander, G. A., Figueroa, R. Organotrifluoroborates: Expanding organoboron chemistry. *Aldrichimica Acta* **2005,** *38* (2), 49-56.

127. Ramage, R., Green, J. N^G-2,2,5,7,8-pentamethylchroman-6-sulphonyl-L-arginine: A new acid labile derivative for peptide synthesis. *Tetrahedron Lett.* **1987,** *28* (20), 2287-2290.

128. Molander, G. A., Ellis, N. Organotrifluoroborates: Protected boronic acids that expand the versatility of the Suzuki coupling reaction. *Acc. Chem. Res.* **2007,** *40* (4), 275-286.

129. Butters, M., Harvey, J. N., Jover, J., Lennox, A. J. J., Lloyd-Jones, G. C., Murray, P. M. Aryl Trifluoroborates in Suzuki–Miyaura Coupling: The Roles of

Endogenous Aryl Boronic Acid and Fluoride. *Angew. Chem. Int. Ed.* **2010,** *49* (30), 5156-5160.

130. McLaughlin, P. A., Verkade, J. G. Fluoride-Catalyzed Reduction of Palladium(II) to Palladium(0)−Phosphine Complexes. *Organometallics* **1998,** *17* (26), 5937-5940.

131. Uno, T., Beausoleil, E., Goldsmith, R. A., Levine, B. H., Zuckermann, R. N. New submonomers for poly N-substituted glycines (peptoids). *Tetrahedron Lett.* **1999,** *40* (8), 1475-1478.

132. Tipson, R. S. On esters of p-toluenesulfonic acid. *J. Org. Chem.* **1944,** *09* (3), 235-241.

133. Swamy, K. C. K., Kumar, N. N. B., Balaraman, E., Kumar, K. Mitsunobu and Related Reactions: Advances and Applications. *Chem. Rev.* **2009,** *109* (6), 2551-2651.

134. Wang, X. J., Sun, X. F., Zhang, L., Xu, Y. B., Krishnamurthy, D., Senanayake, C. H. Noncryogenic I/Br-Mg exchange of aromatic halides bearing sensitive functional groups using i-PrMgCl-bis 2-(N,N-dimethylamino)ethyl ether complexes. *Org. Lett.* **2006,** *8* (2), 305-307.

135. Skaff, O., Jolliffe, K. A., Hutton, C. A. Synthesis of the Side Chain Cross-Linked Tyrosine Oligomers Dityrosine, Trityrosine, and Pulcherosine. *J. Org. Chem.* **2005,** *70* (18), 7353-7363.

136. Osby, J. O., Heinzman, S. W., Ganem, B. Studies on the mechanism of transition-metal-assisted sodium borohydride and lithium aluminum hydride reductions. *J. Am. Chem. Soc.* **1986,** *108* (1), 67-72.

137. Satoh, T., Suzuki, S., Suzuki, Y., Miyaji, Y., Imai, Z. Reduction of organic compounds with sodium borohydride-transition metal salt systems : Reduction of organic nitrile, nitro and amide compounds to primary amines. *Tetrahedron Lett.* **1969,** *10* (52), 4555-4558.

138. Lange, R. G. Cleavage of Alkyl o-Hydroxyphenyl Ethers. *J. Org. Chem.* **1962,** *27* (6), 2037-2039.

139. Chakraborti, A., Kaur, G. One-pot synthesis of nitriles from aldehydes under microwave irradiation: Influence of the medium and mode of microwave irradiation on product formation. *Tetrahedron* **1999,** *55* (46), 13265-13268.

140. Mali, R. S., Papalkar, A. S. Synthesis of naturally occurring cinnamyl cinnamates. *J. Chem. Res., Synop* . **2001,** (10), 433-435.

141. Clayden, J., Greeves, N., Warren, S., Wothers, P., *Organic Chemistry*. Oxford University Press: Oxford, UK, 2001.

142. Evans, C. G., Thomas, J. D. R. Effect of unsaturated substituents on the hydrolysis of esters. *Journal of the Chemical Society B: Physical Organic* **1971**, 1502-1504.

143. Lensink, M. F., Christiaens, B., Vandekerckhove, J., Prochiantz, A., Rosseneu, M. Penetratin-membrane association: W48/R52/W56 shield the peptide from the aqueous phase. *Biophys. J.* **2005,** *88* (2), 939-52.

144. Burley, S. K., Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **1986,** *203* (2), 139-143.

145. Gallivan, J. P., Dougherty, D. A. Cation-pi interactions in structural biology. *Proc. Natl. Acad. Sci. USA* **1999,** *96* (17), 9459-9464.

146. Kumaki, Y., Nitta, K., Hikichi, K., Matsumoto, T., Matsushima, N. Side Chain-Side Chain Interactions of Arginine with Tyrosine and Aspartic Acid in Arg/Gly/Tyr-Rich Domains within Plant Glycine-Rich RNA Binding Proteins. *J. Biochem.* **2004,** *136* (1), 29-37.

147. Feichtinger, K., Sings, H. L., Baker, T. J., Matthews, K., Goodman, M. Triurethane-Protected Guanidines and Triflyldiurethane-Protected Guanidines: New Reagents for Guanidinylation Reactions. *J. Org. Chem.* **1998,** *63* (23), 8432-8439.

148. Kulkarni, A. A., Kingsbury, S. R., Tudzarova, S., Hong, H. K., Loddo, M., Rashid, M., Rodriguez-Acebes, S., Prevost, A. T., Ledermann, J. A., Stoeber, K., Williams, G. H. Cdc7 kinase is a predictor of survival and a novel therapeutic target in epithelial ovarian carcinoma. *Clin. Cancer. Res.* **2009,** *15* (7), 2417-25.

149. Rodriguez-Acebes, S., Proctor, I., Loddo, M., Wollenschlaeger, A., Rashid, M., Falzon, M., Prevost, A. T., Sainsbury, R., Stoeber, K., Williams, G. H. Targeting DNA Replication before it Starts: Cdc7 as a Therapeutic Target in p53-Mutant Breast Cancers. *Am. J. Pathol.* **2010,** *177* (4), 2034-2045.

150. Won, Y.-W., Kim, H. A., Lee, M., Kim, Y.-H. Reducible Poly(oligo-D-arginine) for Enhanced Gene Expression in Mouse Lung by Intratracheal Injection. *Mol. Ther.* **2010,** *18* (4), 734-742.

151. Aubry, S., Burlina, F., Dupont, E., Delaroche, D., Joliot, A., Lavielle, S., Chassaing, G., Sagan, S. Cell-surface thiols affect cell entry of disulfide-conjugated peptides. *FASEB J.* **2009,** *23* (9), 2956-67.

152. Alberts, B., *Molecular Biology of the Cell*. 4th ed.; Garland Science: New York, 2002.

153. Cheng, C. J., Saltzman, W. M. Enhanced siRNA delivery into cells by exploiting the synergy between targeting ligands and cell-penetrating peptides. *Biomaterials* **2011,** *32* (26), 6194-6203.

154. Dunehoo, A. L., Anderson, M., Majumdar, S., Kobayashi, N., Berkland, C., Siahaan, T. J. Cell adhesion molecules for targeted drug delivery. *J. Pharm. Sci.* **2006,** *95* (9), 1856-1872.

155. Schottelius, M., Laufer, B., Kessler, H., Wester, H. J. Ligands for mapping alphavbeta3-integrin expression *in vivo*. *Acc. Chem. Res.* **2009,** *42* (7), 969-80.

156. Mulloy, B., Forster, M. J., Jones, C., Davies, D. B. NMR and molecular-modeling studies of the solution conformation of heparin. *Biochem. J.* **1993,** *293*, 849-858.

157. Lodish, H. F., *Molecular Cell Biology*. 5th ed.; W. H. Freeman: New York, 2003.

158. Cael, J. J., Winter, W. T., Arnott, S. Calcium chondroitin 4-sulfate - molecular-conformation and organization of polysaccharide chains in a proteoglycan. *J. Mol. Biol.* **1978,** *125* (1), 21-42.

159. Carpino, L. A. 1-Hydroxy-7-azabenzotriazole. An efficient peptide coupling additive. *J. Am. Chem. Soc.* **1993,** *115* (10), 4397-4398.

160. Bernardi, D., Ba, L. A., Kirsch, G. First synthesis of N-acylated photoactivatable analogues of glutathione bearing an aryl azide moiety. *Synthesis-Stuttgart* **2007,** (1), 140-144.

161. Krieger, E., Koraimann, G., Vriend, G. Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins* **2002,** *47* (3), 393-402.

162. Wang, J. M., Cieplak, P., Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000,** *21* (12), 1049-1074.

163. Davis, M. E., Zuckerman, J. E., Choi, C. H. J., Seligson, D., Tolcher, A., Alabi, C. A., Yen, Y., Heidel, J. D., Ribas, A. Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature* **2010,** *464* (7291), 1067-1070.

164. Wang, Y., Li, Z., Han, Y., Liang, L. H., Ji, A. Nanoparticle-based delivery system for application of siRNA *in vivo*. *Curr Drug Metab* **2010,** *11* (2), 182-96.

165. Vives, E., Schmidt, J., Pelegrin, A. Cell-penetrating and cell-targeting peptides in drug delivery. *Biochim. Biophys. Acta* **2008,** *1786* (2), 126-38.

Appendix I

Publication

Appendix I - Publication

The following paper based on the work in this thesis was accepted for publication in the journal Chemical Biology and Drug Design on 24th September 2011.

Exploring the interaction between siRNA and the SMoC biomolecule transporters: implications for small molecule mediated delivery of siRNA.

Matt Gooding[1], Slavica Tudzarova[1], Roberta J. Worthington[2], Michela I. Simone[3], Dimitris Lagos, Anne-Sophie Rebstock[1], Cristina Visintin[1], Sarah R. Kingsbury[1], Heike Laman[4], Chris Boshoff[5], Gareth H. Williams[1], Kai Stoeber[1], and David L. Selwood[1].

[1] The Wolfson Institute for Biomedical Research, UCL, Gower Street, London, WC1E 6BT, UK.

[2] Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, USA.

[3] School of Chemistry, University of Sydney, Sydney, NSW 2006, Australia.

[4] University of Cambridge, Department of Pathology, Division of Cellular and Genetic Pathology, Tennis Court Road, Cambridge CB2 1QP, UK.

[5] UCL Cancer Institute, UCL, London, WC1E 6BT, UK.

*Corresponding author: d.selwood@ucl.ac.uk

Abstract

The Small Molecule Carrier (SMoC) class of biomolecule transporters, modelled on the third helix of the Antennapedia homeodomain, has previously been shown to transport active proteins into cells. Here, we show an improved synthetic route to SMoCs, including Molander chemistry using trifluoroborate salts to improve the yield of the Suzuki-Miyaura coupling step for formation of the biphenyl backbone. The required boronic acids could be formed by reaction of a 2-(dimethylamino)ethyl ether-modified aryl Grignard reagent with triisopropyl borate. The potential for the use of SMoCs as oligonucleotide transporting agents was also explored by characterising the interactions between SMoCs and siRNA. Molecular dynamics and NMR analysis indicated that the SMoC guanidines are stabilised by π-cation

interactions with the biphenyl system, thus not only increasing the basicity or pKa but also shielding the charge. The binding affinities of various SMoCs for siRNA were investigated using ITC and gel shift assays. SMoC-mediated siRNA delivery to cultured fibroblasts is demonstrated, showing that SMoCs possess the ability to transport functional siRNA into cells. Knockdown of Cdc7 kinase, a target for cancer, is achieved.

Introduction

The use of short interfering RNA (siRNA) to knockdown target genes via the mechanism of RNA interference (RNAi) has received much attention over the last decade due to the wide-ranging therapeutic implications of this technology. Indeed, the first clinical trials of siRNA-based drugs have started to emerge in recent years, with many more potential targets having been identified both *in vitro* and *in vivo* (1, 2). There are, however, major challenges to the development of RNAi therapeutics, including degradation, specificity and delivery. In order to knockdown its target gene, the siRNA must enter the cell and bind to a protein complex called the RNA-induced silencing complex (RISC) (3). siRNA molecules are large and highly negatively charged, rendering them unable to cross the lipid bilayer of the cell membrane. Therefore, the development of a safe, effective delivery agent for siRNA remains one of the biggest challenges in biological therapeutics.

Several types of delivery vectors have been successfully used to transport RNA into cells, including liposomes (4), cationic polymers (5), and cell penetrating peptides (CPPs) (6-9). Among the CPPs that have been shown to be successful in delivering siRNA are TAT, penetratin, transportan and poly-arginine (8). Amphipathic CPPs, which display both cationic and lipophilic residues in the same molecule, have been shown to be particularly effective delivery agents, and several successful synthetic amphipathic CPPs have been designed such as CADY (7, 10) and MPG (9). Although the mechanism for CPP transduction is yet to be fully revealed, it is thought that crucial steps include binding to negatively charged cell-surface proteoglycans (7, 11, 12), followed by endocytotic internalisation, translocation (12) and endosomal escape (13). The mechanism for internalization (14) has been widely debated, recent studies point to multiple mechanisms operating (12) (15) with peptide structure playing a key role.

π-Cation interactions were first described by Burley and Petsko in 1986 when they observed that the side chains of cationic amino acids such as lysine and arginine are preferentially located near the π-electron cloud of aromatic amino acids such as phenylalanine, tyrosine and tryptophan (16). It has been found that π-cation

interactions are very common in proteins, with one interaction per 77 residues on average. Arginine residues are more favoured for π-cation interactions than lysine, probably due to the guanidine group on arginine being larger and less solvated, resulting in greater van der Waals interactions with aromatic rings than that of lysine, as well as the ability of guanidines to form hydrogen bonds and π-cation interactions simultaneously, whereas the lysine sidechain can only participate in one type of binding at a time (17).

Previously, we reported on a class of small molecule carriers (SMoCs) based on amphipathic CPPs which were capable of delivering functional proteins into cells (Figure 1) (18). We hypothesize that since SMoCs mimic guanidine-rich CPPs, possessing guanidinium cations linked to a lipophilic biphenyl system (19), they may be able to internalise siRNA in the same way as amphipathic CPPs such as MPG. The positively charged guanidinium cations also enable the formation of electrostatic complexes with siRNA as have been observed for several cationic peptides (6, 10, 20). In this paper, we study the potential for SMoCs to form non-covalent complexes with siRNA by investigating the pKa of the SMoC guanidine groups, as well as measuring the affinity of different SMoCs for siRNA. We also examine the intramolecular structure of the SMoCs, and provide evidence to indicate that the guanidine groups are stabilised by π-cation interactions (Figure 1) (17). Moreover we explore whether these simple SMoCs may be used as siRNA delivery agents *in vitro*.

Materials and methods

Synthetic Chemistry

Starting materials were either commercially available or synthesized according to methods reported in the literature. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker AMX-300 or a Bruker AMX-500 spectrometer. Chemical shifts are reported as ppm relative to TMS internal standard. Mass spectra were recorded on a Fisons VG70-SE spectrometer (EI, FAB) or an Agilent 6100 Series LC-mass spectrometer using a C-18 or C-4 columns. Microwave reactions were carried out using a CEM Discover microwave. Polyarginine peptides were obtained from Peptide Protein Research, Hampshire, UK. siRNA sequences were purchased from Dharmacon RNAi Technologies. The sequences used were GAPDH sense 5' UGG UUU ACA UGU UCC AAU AUU 3', antisense 5' Phosphate U AUU GGA ACA UGU AAA ACC UU 3'; *CDC7* sense 5' GCU CAG CAG GAA AGG UGU UUU 3', antisense 5' AAC ACC UUU CCU GCU GAG CUU.

Molecular Dynamics

Molecular dynamics simulations were run for compounds **14**, **18**, **19** and **20** in the YASARA (21) software and using the AMBER99 force field with default parameters. The starting structures were contained in a simulation cell which extended 5 Å larger than the SMoC structure along each axis *in vacuo*, the default pH set to 7.0 and the structures were minimised under the AMBER99 force field (22) starting at 298 K for 1 ps. The simulation substep time was set to 1 fs and the intramolecular forces were recalculated every 2 substeps and the temperature controlled by velocity rescaling. A 5 ns simulation was performed for each structure, after which snapshots were taken and analysed in the DS Visualizer software. No atoms were kept frozen during the simulations.

NMR study

The SMoC compounds were dissolved in 0.5 mL deuterium oxide to a final concentration of around 10 mg/ml. 100 mM NaOH was added in small aliquots (5-30 µl) and the pH of the solution measured in the NMR tube using a micro electrode attached to a pH meter. NMR spectra were recorded using a 400 MHz Bruker spectrometer at 20°C.

Gel shifts

1.7 nmol GAPDH siRNA (Dharmacon RNAi Technologies) was mixed in RNase-free water with varying concentrations of either SMoC, tetraarginine or octaarginine in the molar ratios of 1/1, 1/5, 1/10, 1/20 and 1/50. The complexes were incubated for 30 min at room temperature. The samples were diluted with loading buffer containing 95% formaldehyde, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF, 0.025% ethidium bromide and 0.5 mM EDTA (Fermentas UK), and analysed by gel electrophoresis on a 1% agarose gel in a running buffer containing 6.7 % formaldehyde, 0.04 M 3-morphilinopropane-1-sulfonic acid (MOPS), 0.01 M sodium acetate and 0.001 M EDTA. The gels were analysed by UV illumination and the ethidium bromide bands quantified using the ImageJ software. The EtBr intensities were plotted, a dose response curve fitted and EC_{50} values calculated using the Origin software.

Isothermal calorimetry (ITC)

6G-SMoC at a concentration of 250 µM was loaded into the syringe of a Microcal VP-ITC calorimeter (450 µl), and the cell (1.8 ml) filled with human GAPDH siRNA (Dharmacon RNAi Technologies) at a concentration of 3.5 µM. A total of 37 injections of 8 µl each were made at 4 min intervals to ensure total saturation of the siRNA. The binding curve was plotted using Origin 6.0 (Microcal) and the binding constants calculated.

Cell culture and cell cycle analysis

IMR90 (ATCC# CCL-186), a diploid primary human fibroblast adherent cell line derived from fetal lung tissue, was obtained from LGC Standards (Middlesex, UK) at population doubling (PD) 12. All culture passages and population doublings were recorded, and experiments were performed with IMR90 cells under a PD of 22. IMR90 cells were cultured at 37 °C with 5% CO_2 in DMEM (Invitrogen, Paisley, UK) supplemented with 10% defined FCS (Invitrogen), 100U/ml penicillin and 100 μg/ml streptomycin.

Immunoblotting

For western blot analysis, cells were harvested and whole cell extract (WCE) was prepared in modified RIPA lysis buffer (50 mM Tris-HCl, 300 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM EDTA and protease inhibitors). After sonication of lysates for 10 sec protein concentration was determined using the DC Bio-Rad protein assay kit (Bio-Rad Hemel Hempstead, UK). 40 μg of total protein was loaded in each lane and separated by 4% to 20% SDS-PAGE. Protein was transferred from polyacrylamide gels onto PVDF membranes (Bio-Rad) by semi-dry electroblotting. Blocking, antibody incubations and washing steps were performed as described (23). Antibodies used for immunoblotting included: $p21^{WAF1}$ from BD Biosciences (Oxford, UK), p53 (Ab-6) from Merck (Beeston, UK); Cdc7 from MBL International (Woburn, MA), and β-actin from Sigma (Gillingham, UK); Mcm2 phospho-Ser-40/41 from Bethyl Laboratories (Montgomery, TX). Affinity-purified rabbit polyclonal antibodies for ASK were generated by Eurogentec (Seraing, Belgium) following the manufacturer's protocol.

Immunofluorescence

The transfection efficiency of SMoC compared to Lipofectamine 2000 was determined with fluorescein-conjugated nonspecific siRNA-transfected cells (FO) (BLOCK-iT Transfection Optimization kit; Invitrogen) using a Leica TCS SP confocal fluorescence microscope. IMR-90 fibroblasts grown on coverslips were treated for 24 h with complexes of 4G SMoC (40 μg/ml, 35 μM) and either 10 nM FO, 100 nM FO or 100 nM FO without SMoC. After the treatment cells were washed in PBS and fixed in 3.7 % paraformaldehyde. Following fixation, cells were treated with 0.1% Triton X in PBS, blocked with 1% BSA in PBS for 1h and mounted directly, or first stained with AlexaFluor 546 Phalloidin according to manufacturer's

instructions (Invitrogen) and then mounted in Vectashield mounting medium (Vector Laboratories, Peterborough, UK) with 1.5 μg/ml DAPI to visualize DNA.

RNA interference

CDC7 expression was inhibited with double-stranded RNA oligo for CDC7 synthesized by Ambion (Warrington, UK) in complex with Lipofectamine 2000 (Invitrogen) as described (24) or with 4G SMoC. Briefly, IMR90 cells were seeded at a density to reach 50% confluence on the day of transfection. The transient transfections were performed using 10 nM of CDC7 siRNA duplex for 72h. siRNAs were complexed with transfection reagent in serum-free and antibiotic-free culture medium. Selective silencing of the corresponding gene was confirmed by qRT-PCR and western blotting.

RNA extraction and qRT-PCR

To evaluate the efficiency of transfection with CDC7 siRNA using SMoC compared to Lipofectamine 2000, CDC7 mRNA levels were determined by qRT-PCR as described (24). Briefly, total RNA was isolated using a PureLink Micro-to-Midi kit (Invitrogen) according to the manufacturer's instructions. Reverse transcription reactions using 40 ng of total RNA in a final reaction volume of 20 μl were performed in one step using SuperScript III Platinum SYBR Green One Step qRT-PCR Kit (Invitrogen). Relative quantitation data were obtained using the comparative C_t method with Realplex software according to the manufacturer's protocol (Eppendorf, Heidelberg, Germany). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used to normalize each of the extracts for amplifiable human DNA. Primers were provided by Eurofins MWG Operon (Ebersberg, Germany). Cycle conditions are available upon request.

Results and Discussion

Improved SMoC synthetic route to 4G-SMoC

Previously, we described the synthesis of a phenyl boronic ester which acts as a key intermediate in the SMoC synthetic pathway (18). The boronic ester takes part in a Suzuki-Miyaura coupling which generates the biphenyl backbone of the SMoC structure, linked via ether bonds to Z-protected amine groups. As an alternative to the boronic ester, we investigated the formation of the equivalent boronic acid from the aryl bromide **1** using a Grignard reagent with 2-(dimethylamino)ethyl ether

followed by quenching with trimethyl borate to yield the boronic acid **2** (Figure 2). As described by Wang *et al.* (25), the organic ligand moderates the reactivity of the aryl Grignard reagent in this reaction, reducing possible side reactions. Careful control of the reaction temperature was required to obtain a synthetically useful yield of 40%.

In addition, the potassium trifluoroborate salt was synthesized (Figure 3). Trifluoroborate salts, developed by Molander *et al* (26, 27), are extremely stable to air and moisture, allowing large scale synthesis of this key intermediate for long-term storage. The pinacolboronate ester **3** synthesized previously was treated with potassium bifluoride in methanol, resulting in an 88% yield of the salt **4** which is precipitated from the reaction mixture and is easily isolated by filtration and washing with organic solvents. Suzuki reactions with this salt and the alkylated nitrile **5** gave cleaner conversion to the biphenyl **6** with less of the deboronated side product that was difficult to separate from the desired product.

A problematic step in the original SMoC synthesis was the addition of Boc-protected guanidine groups using the expensive *N*,*N*-di-Boc-*N'*-trifluoromethanesulfonyl-guanidine reagent, which resulted in modest yields. To improve this step, a less expensive and more stable guanidinylation reagent was used, *N*,*N'*-Di-Boc-1*H*-pyrazole-1-carboxamidine. Thus following removal of the CBz groups using HBr in acetic acid, the use of this reagent increased the yield of guanidinylation from ~65% to over 70% of **7** at a much reduced cost of materials.

Finally, the reduction of the nitrile **7** in order to produce a primary amine for attachment of diverse linkers was another low-yielding step from our previous route. The previous synthesis which used Raney Nickel/H_2 resulted in 40-50% yields. Reduction using sodium borohydride catalysed by cobalt chloride increased the yield of this step to ~60% of **8** (28, 29). Amide coupling with the thiopyridylpropionic acid O-succinimide ester gave the desired thiopyridyl intermediate **9** and deprotection with TFA provided **10**. Synthesis of the disulfide **11** was easily accomplished in a two step procedure with DTT and then aerial oxidation.

The final optimised synthesis of 4G-SMoC is therefore shown in Figure 4. This improved method has enabled us for the first time to synthesize 4G-SMoC on a >200 mg scale.

Synthesis of simple SMoC analogs

The simple 1G- or 2G-SMoC derivatives were synthesized (Figure 5) using similar methods to those described above. Thus simple alkylation utilising the Cbz protected mesylate **12** gave access to the simple monophenyl **13**, and biphenyl **15-17** intermediates. Guanidylation and deprotection utilising the new reagent method above gave **14** and **18-20**. These compounds were synthesized from commercially available mono- or biphenyl starting materials, eliminating the Suzuki coupling step.

6G-SMoC terphenyl analogue synthesis

In order to maximise binding to siRNA, a new SMoC was synthesized with six guanidine groups attached to a triphenyl ring structure. The 6G-SMoC was synthesized (Figure 6) starting from the alkylated bromoiodocatechol synthesized previously (18) via coupling with the boronic ester, separation of the mono-**22** and di-coupled product **21** and second coupling of the mono-coupled product. This was followed by deprotection of the Z protecting groups and subsequent guanidinylation (30) to give **23**. Final removal of the Boc groups gave **24**.

Molecular Dynamics

In order to gain more information about the structural conformation of the SMoC compounds, some simple SMoC-like structures were analysed using the molecular dynamics software YASARA under the AMBER99 force field (Figure 7). It was found that in the case of a 2G-SMoC, possessing two guanidine side chains at the *ortho* position to the biphenyl bond, the guanidine groups form 'T-shaped' π-cation interactions with the opposite ring. This π-cation interaction is also seen in a 1G-SMoC structure with one guanidine side chain *ortho* to the biphenyl bond. However, when the side chain is moved to the *meta* position, the carbon chain is not long enough to allow interaction with the opposite ring, and a weaker 'parallel' π-cation interaction is seen between the guanidine and the adjacent ring. When only a single phenyl ring is present, no interaction occurs, suggesting that the electron density of one ring is insufficient to promote a π-cation interaction. These results suggest that the SMoC guanidinium ions may be stabilised by the presence of an additional phenyl ring (Figure 1), thus allowing stronger electrostatic interactions with siRNA. A movie of the molecular dynamics simulation is presented in the supporting information.

NMR pKa Study

In order to investigate the strength of the proposed π-cation interaction, NMR was used to determine the pKa of SMoC guanidine groups by directly measuring the change in chemical shift of the adjacent methylene hydrogens (31, 32). A change in the pKa of the guanidine would be expected if a π-cation interaction was present (31).

For each compound, the change in ionisation state of the guanidine groups was measured using NMR by titrating each compound with sodium hydroxide and recording the ^{1}H chemical shifts of the CH_2 protons adjacent to the guanidine group against the pH of the solution (Figure 8). The pKa of the guanidine group is given by the Henderson-Hasselbalch equation (33):

$$pKa = pH + \log\frac{[BH+]}{[B]}$$

Deprotonation of the guanidine is expected to have a significant effect on the chemical shift of the protons on the adjacent methylene group due to the change in electron density around the nitrogen atom, resulting in increased shielding of the CH_2 protons. The chemical shift of the charged form (δB^{+}) may be determined by measuring the chemical shift at a low pH. For these compounds, the chemical shift was found to be at its maximum below pH 10. Likewise, at very high pH (pH 14) the chemical shift of the deprotonated guanidine groups may be measured (δB). Thus, the chemical shift at a given pH (δA) is representative of the proportion of groups that are protonated at a given time. Therefore, the equation may be expressed in terms of change in chemical shift:

$$pKa = pH + \log\frac{\delta A - \delta B}{\delta B^{+} - \delta A}$$

Hence, by plotting the log of the ratio of the charged and uncharged forms expressed as chemical shifts against the pH of the solution (Figure 9), the pKa may be calculated from the y-intercept of the linear trend line.

The results of these experiments (Table 1) show that the 2G-SMoC with the bi-phenyl substituted at the ortho position to the biphenyl bond, **18**, has the highest pKa (13.33), suggesting that the guanidine groups in this compound benefit from the greatest stabilisation. The mono-phenyl compound, **14** (12.86), has the lowest pKa

and the closest to free arginine (12.48) (34), suggesting very little stabilisation is available to the guanidinium ion. These results also show that substitution position is also important for stabilisation, since the meta-substituted compound, **19**, has a lower pKa (13.09) than the ortho-substituted **20** (13.26). These results suggest that the biphenyl system is able to provide some stabilisation to the positively charged guanidinium ion. The energy of this stabilisation may be calculated by comparing the pKa values of the mono-phenyl compound, **20**, and the mono-substituted biphenyl compound **19** using the equation:

$$\Delta\Delta G^{0}_{pKa} = 2.303RT\Delta pKa$$

This gives a free energy of stabilisation of -2.3 kJ mol^{-1} for each guanidine group compared to the mono-phenyl compound, for which it was assumed that there is very little π-cation stabilisation, due to its low pKa. This value is the increase in the free energy of the guanidine group gained by the addition of a phenyl ring to the mono-phenyl compound **14**.

ITC of 6G-SMoC indicates a high enthalpic contribution to binding.

In order to calculate the strength of the interactions between SMoC and siRNA, isothermal titration calorimetry (ITC) was used to measure the heat evolved when small aliquots of 6G-SMoC, **24** were injected into a solution of a standard 21-bp siRNA (Figure 10). A binding constant (K) of 87.3 μM was calculated with a binding stoichiometry of 4.4:1 (SMoC:siRNA). Binding was dominated by enthalpic contributions (-1646 Kcal/mol) with a minor or negligible entropic contribution. This is consistent with electrostatic (charge) interactions between the guanidines and the phosphate groups on the siRNA dominating binding.

Gel shift studies on siRNA show that SMoCs can effectively complex to siRNA duplexes.

The interactions of 4G- and 6G-SMoC compounds with siRNA were analysed by gel electrophoresis of SMoC/siRNA complexes at different molar ratios. Tetra- and octa-arginine peptides were also tested in order to establish the relative binding affinity of SMoCs compared to these established cell delivery peptides. As shown in Figure 11, all SMoCs and peptides were able to fully shield the siRNA negative charges at ratios less than 20:1. Since the observable band on the gel is proportional to the free siRNA concentration we utilised this value to provide a convenient graphical representation of the relative affinities and to calculate the approximate binding

constants of the analogues. The ethidium bromide (EtBr) fluorescence of each band was quantified and the data plotted as dose response curves (Figure 12). 6G-SMoC, **24** had a slightly lower binding affinity than 4G SMoC, **10** (EC_{50} of 110 μM compared to 70 μM) and 4G-SMoC had a higher affinity than R4 (EC_{50} of 70 μM compared to 120 μM). However R8 had a higher affinity (3 μM) than any of the SMoCs. The disulfide variant $(4G\text{-}SMoCS)_2$, **11** has the same binding affinity as the monomer form, probably due to reduction of the disulfide bond in the gel running buffer. 4G-SMoC, **10** binds with a higher affinity than the polyarginine peptide R4, indicating an ability to complex to siRNA more efficiently or to shield the negative charges more effectively. This may be attributed to the increased pKa of the SMoC guanidine groups as a result of the π-cation effect described previously.

Confocal analysis of SMoC mediated siRNA uptake

To determine if siRNA-SMoC complexes are taken up by cells, the simple SMoCs described above were complexed with a fluorescently-tagged siRNA oligo at concentrations of either 10 nM or 100 nM and incubated with IMR-90 cells, a diploid human fibroblast adherent cell strain derived from foetal lung tissue, for 24 h. Upon examination under confocal microscopy, only the 4G-SMoC compound **10** showed significant uptake (others not shown). For compound **10**, the fluorescent oligo was observed in the cytoplasm in a punctate distribution for both concentrations, showing that uptake occurs into vesicles, possibly endosomes formed via an endocytotic pathway (Figure 13).

siRNA-SMoC gene and protein knockdown

To determine if SMoCs are capable of transporting siRNA into cells, IMR-90 cells were transfected with SMoC or Lipofectamine complexed with a double-stranded RNA oligo targeted at Cdc7, a cell cycle kinase which has been identified as a new cancer therapeutic target (35). Cancer cells establish only a limited number of DNA replication forks under Cdc7 rate-limiting conditions, causing fork stalling and collapse during an abortive S phase and triggering apoptotic cell death. Tumour cell specificity comes from normal cells avoiding lethal S phase progression in the presence of low Cdc7 kinase levels by eliciting a p53-dependent replication origin activation checkpoint response that arrests the cells at the G1/S boundary (24). Seventy-two hours after the treatment, IMR90 cells were harvested, and CDC7 mRNA knockdown was monitored by qRT-PCR. For each compound tested, the CDC7 mRNA level was determined relative to control cells treated only with Lipofectamine or SMoC. Treatment with complexes of 4G-SMoC, **10** and CDC7-siRNA reduced CDC7 mRNA levels by 61%. These results were similar to the CDC7 knockdown achieved with Lipofectamine plus CDC7-siRNA, which reduced

mRNA levels by 80% (Figure 14A). RNAi against CDC7, mediated through 4G-SMoC or Lipofectamine, caused an accumulation of IMR90 cells with G1 DNA content and a concomitant reduction in the S phase fraction, while the fraction of cells with less than 2C DNA content was negligible, indicating that the Cdc7-depleted cells remained viable (Figure 14B). We recently demonstrated the triggering of an origin activation checkpoint in Cdc7-depleted IMR90 cells (24). Consistent with this study, western blotting of extracts prepared from IMR90 cells treated with complexes of 4G-SMoC, **10** and CDC7-siRNA showed that phosphorylation of the replication initiation factor Mcm2 at the Cdc7-dependent phosphor-site Ser-40/41 was markedly reduced when Cdc7 kinase and its regulatory subunit ASK were downregulated. In keeping with activation of the origin activation checkpoint in the Cdc7-depleted cells, we also noted p53 stabilization and increased levels of the CDK inhibitor protein p21 (Figure 14C). Taken together, these data demonstrate that SMoCs are capable of mediating the uptake of siRNAs into cultured human cells without altering their biological activity. Note that 6G-SMoC, **24** was also tested for siRNA transfection but did not result in significant knockdown (data not shown).

Conclusions

The SMoC class of compounds, modelled on amphipathic CPPs, have previously been shown to transport active proteins as well as conjugated dyes into a variety of cell types (18). Here we have improved the synthetic route to the SMoC compounds, including the use of Molander chemistry in the form of a stable trifluoroborate intermediate which can be used to easily build new, varied SMoC-like structures, as well as optimisation of some of the previously difficult steps in order to increase the scale of the synthesis.

We have also shown in this study that SMoCs show potential as siRNA transfection agents. One surprise from this study is that the most effective delivery and knockdown was achieved with the 4G-SMoC, compound **10** where the thiopyridyl group is commonly regarded as a reactive protein thiol reagent. An improvement in CPP delivery has been noted for some cysteine containing peptides (36) where dimerization or oligimerization is indicated. In addition enhancement of internalization has been noted for CPPs containing cysteines conjugated with the 2-nitropyridyl group (37). An interaction with cell surface thiols was suggested for this observation and we speculate that compound **10** may also interact with cell surface thiols to promote internalization.

siRNA delivery remains a significant challenge in biological therapeutics, and an effective transfection agent will open up a vast array of possibilities in the treatment of many diseases by specifically targeting the underlying genetic malfunctions. The

standard in vitro transfection agent, lipofectamine is demonstrably toxic, by several measures (38) indicating the need for further research in this area. Cell penetrating peptides show a great deal of promise as transfection agents, and many studies have successfully introduced siRNA into cells using both covalently- and non-covalently linked CPPs (8). As CPP mimics, it is therefore likely that SMoCs will also possess this ability to act as siRNA delivery vectors, possessing positively charged guanidine groups which allow the formation of electrostatic complexes with siRNA, and are also essential for successful passage across the cell membrane. Our MD and NMR studies indicate that the guanidinium ions on the SMoC side chains are stabilised by π-cation. The potential impact of π-cation effects has also been advanced for the Arg-Trp interaction in the penetratin CPP (39). We probed the affinity of the SMoC-siRNA interaction using two methods ITC (the gold standard for evaluating binding interactions) and simple gel shift analysis. These were in good agreement indicating binding affinities in the range 70 to 110 μM. The poly arginine peptide R8 had an affinity of 30 μM. The optimal binding affinity for SMoC-siRNA delivery remains to be determined. As we show here, SMoCs can deliver biologically active siRNA into cultured human cells, and achieved RNAi knockdown with a similar magnitude to the widely used Lipofectamine reagent. Further optimisation of SMoCs to increase the uptake of SMoC-siRNA complexes will be reported in due course.

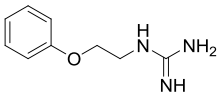
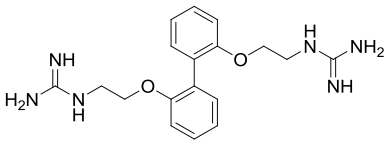
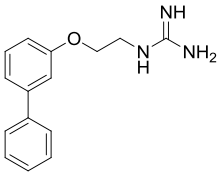
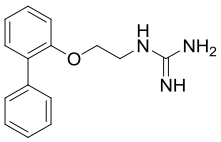
Acknowledgements

We thank the Association for International Cancer Research (AICR) for funding this project (Grant Ref: 06-0076). This work was also partly funded by Cancer Research UK Scientific Programme Grant C428/A6263 (KS and GHW). We would also like to thank Tina Divita and Matt Webster of the ISMB Biophysics Centre at Birkbeck College for their assistance with the ITC work.

Supporting information

Full experimental details of chemistry syntheses and a movie of the molecular dynamics simulation are presented.

Table 1. pKa values calculated from the y-intercepts from Figure .

Compound No.	Structure	pKa
14		12.86
18		13.33
19		13.09
20		13.26

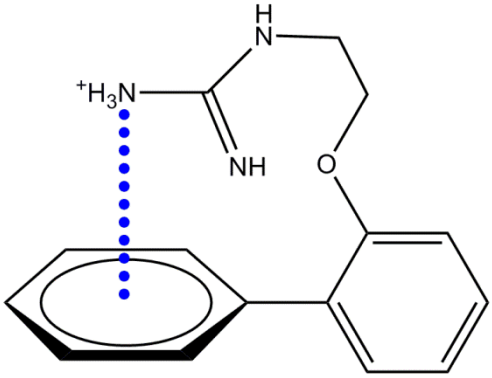


Figure 1. Illustration of the π-cation effect for a biphenyl SMoC molecule.

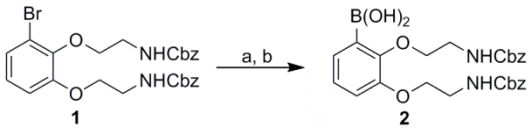


Figure 2. Boronic acid synthesis. Reagents: a) bis[2-(*N*,*N*-dimethylamino)ethyl] ether, ^{i}PrMgCl, THF, 10-15°C, 24 h; b) Trimethyl borate, THF, room temperature, 1 h (40%).

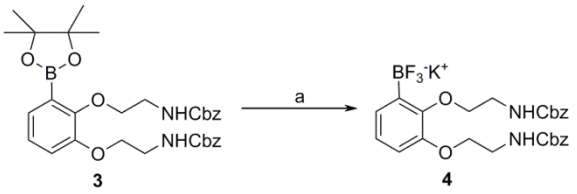


Figure 3. Conversion of boronic ester to trifluoroborate salt. Reagents: a) Aqueous potassium bifluoride, MeOH, room temperature, 3 h (88%).

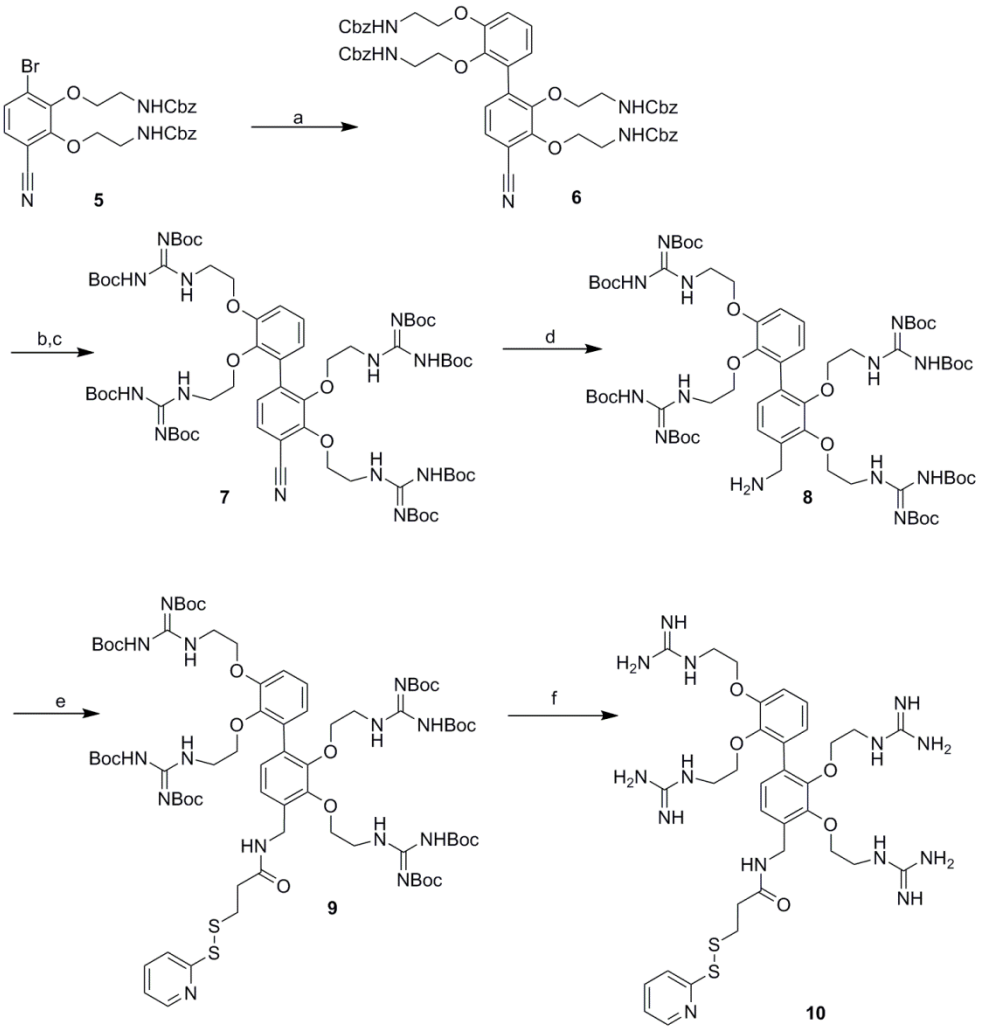


Figure 4. New synthetic route to 4G-SMoC. Reagents: a) **4**, $PdCl_2dppf.CH_2Cl_2$, Et_3N, iPrOH/H_2O, 82°C, 18 h (79%); b) 30% HBr in acetic acid, DCM, room temperature, 1 h; c) N,N'-Di-Boc-1H-pyrazole-1-carboxamidine, DIEA, DCM, room temperature, 18 h (71% over two steps); d) $CoCl_2.6H_2O$, $NaBH_4$, MeOH, room temperature, 1 h (60%); e) 2,5-dioxopyrrolidine-1-yl-3-(pyridine-2-yldisulfanyl)propanoate, DIEA, DCM, room temperature, 18 h (57%); f) TFA/m-

cresol, DCM, room temperature, 3 h (82%); g) Dithiothreitol, H_2O; h) ammonium acetate, air.

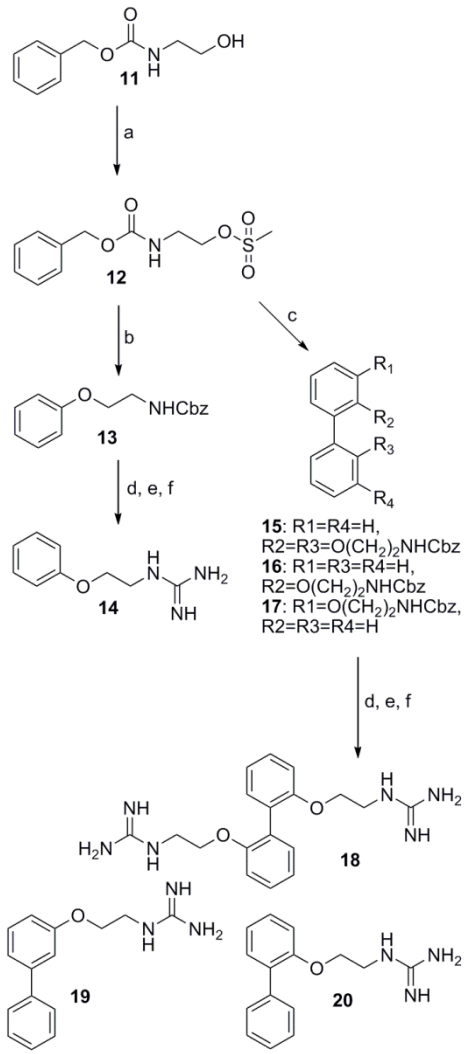


Figure 5. Synthesis of simple SMoC compounds for determination of guanidine pKa. Reagents: a) Phenol, Cs_2CO_3, acetone, 17 h, 65 °C (70%); b) 2-,3-, or 2,2'-hydroxybiphenol, Cs_2CO_3, acetone, 17h, 65 °C (**15**: 35%; **16**: 93%; **17**: 100%); c) 30% HBr in acetic acid, DCM, 3 h, room temperature; d) N,N′-Di-Boc-1H-pyrazole-1-carboxamidine, DIPEA, MeCN, 48 h, room temperature (**13a**: 45%; **15a**: 51%; **16a**: 26%; **17a**: 40%); e) TFA/TIPS/H_2O, 4 h, room temperature (100%).

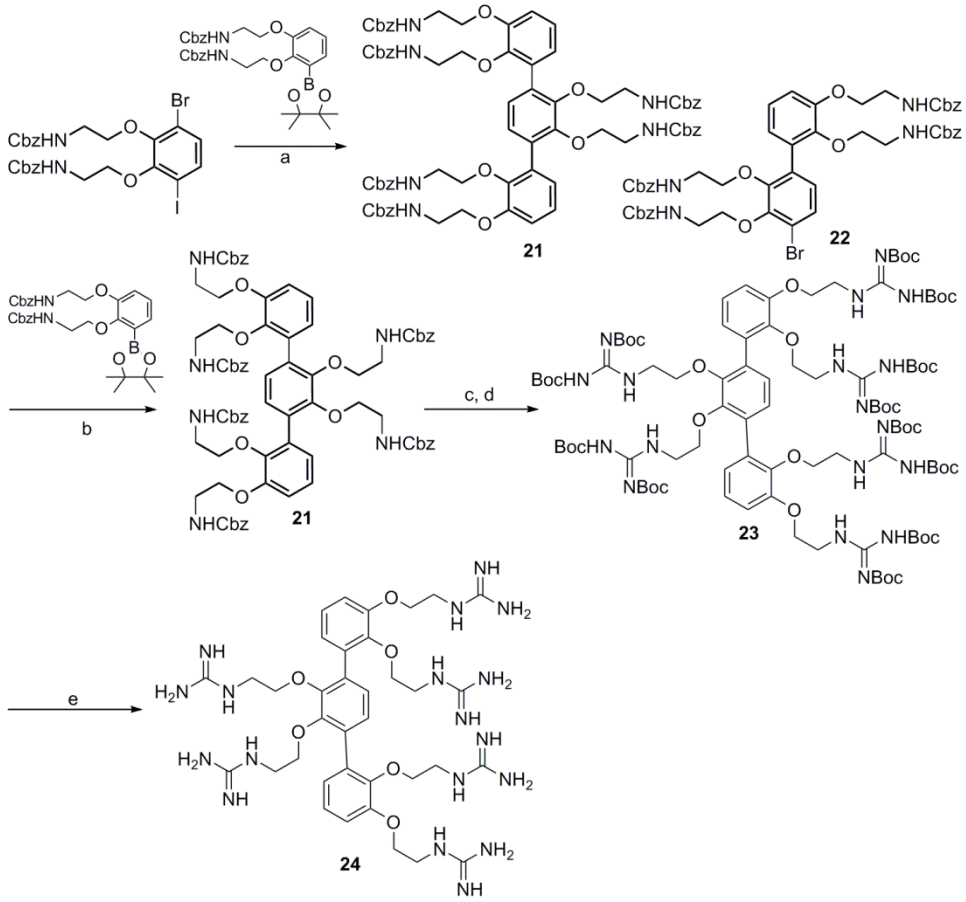


Figure 6. 6G-SMoC synthesis. Reagents: a) $PdCl_2dppf$, K_3PO_4, toluene, water, 100 °C 3 h ; b) $PdCl_2dppf$, K_3PO_4, toluene, water, 100 °C overnight (48% over two steps); c) 30% HBr in acetic acid, DCM, room temperature; d) N, N-di-boc-N'-trifluoromethanesulfonyl-guanidine, DIPEA, DCM, room temperature, overnight (54% over two steps); e) TFA/TIPS/H_2O, 4 h, room temperature (100%).

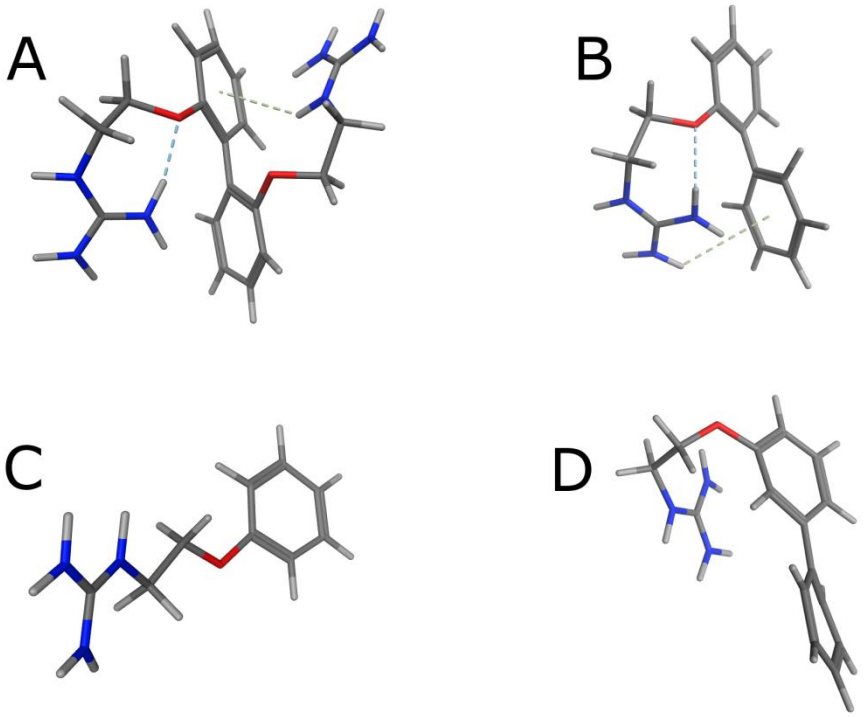


Figure 7. **Snapshots of a molecular dynamics simulation of simple SMoC-like molecules containing one or two guanidine side chains.** In a simple 2G-SMoC molecule (A), T-shaped π-cation interactions form between the side chains and their opposite rings. This is also the case when one side chain is removed (B). For a mono-phenyl structure with one side chain (C), no interaction is present, whereas the addition of a second ring *meta* to the side chain (D) promotes a parallel π-cation interaction onto the adjacent ring.

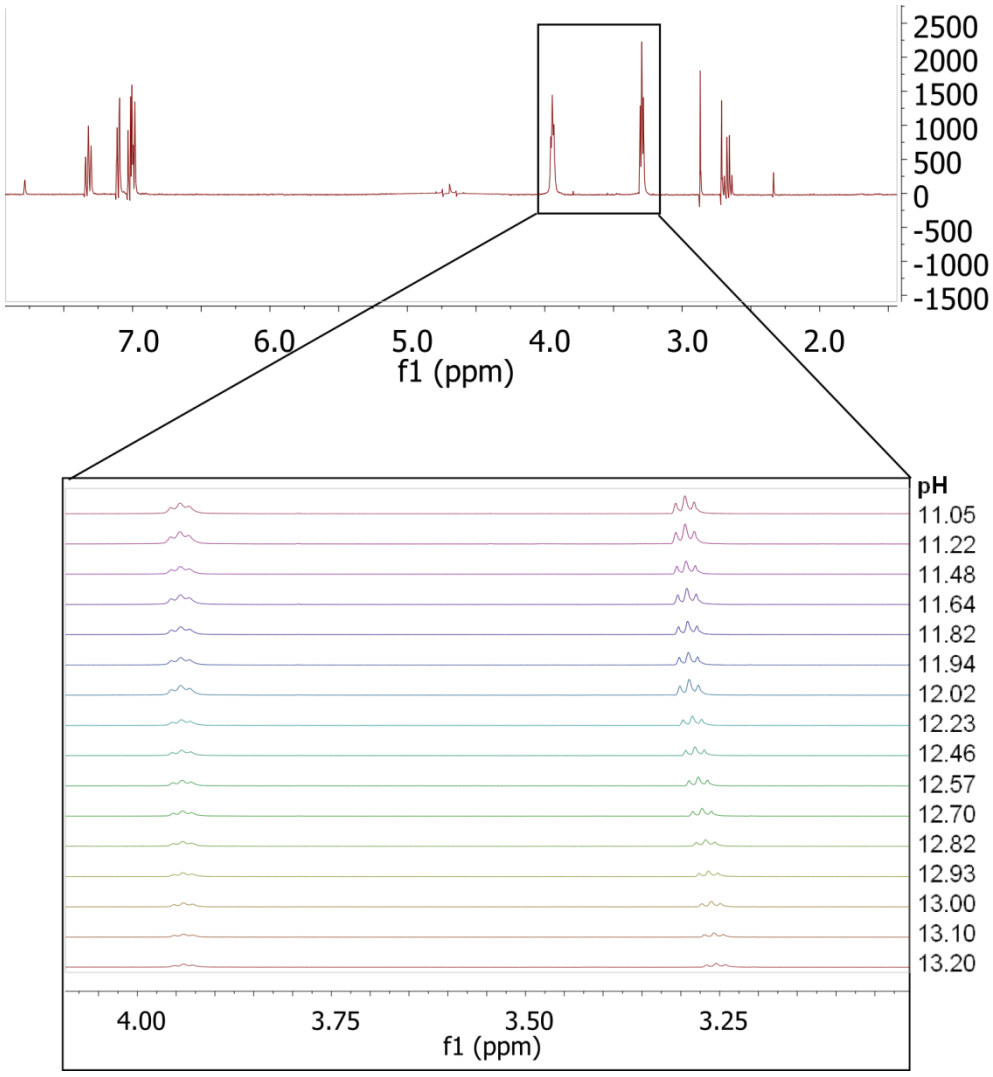


Figure 8. **Change in SMoC NMR spectrum as the pH of the solution is increased.** Expanded section of NMR spectrum of **18** showing the two triplets corresponding to the four ethyl protons at varying pH values. The CH_2 adjacent to the guanidine group (peak at ~3.3 ppm) becomes less shielded at higher pH values, resulting in a shift in the NMR peak, whereas the other peak (3.97) remains static.

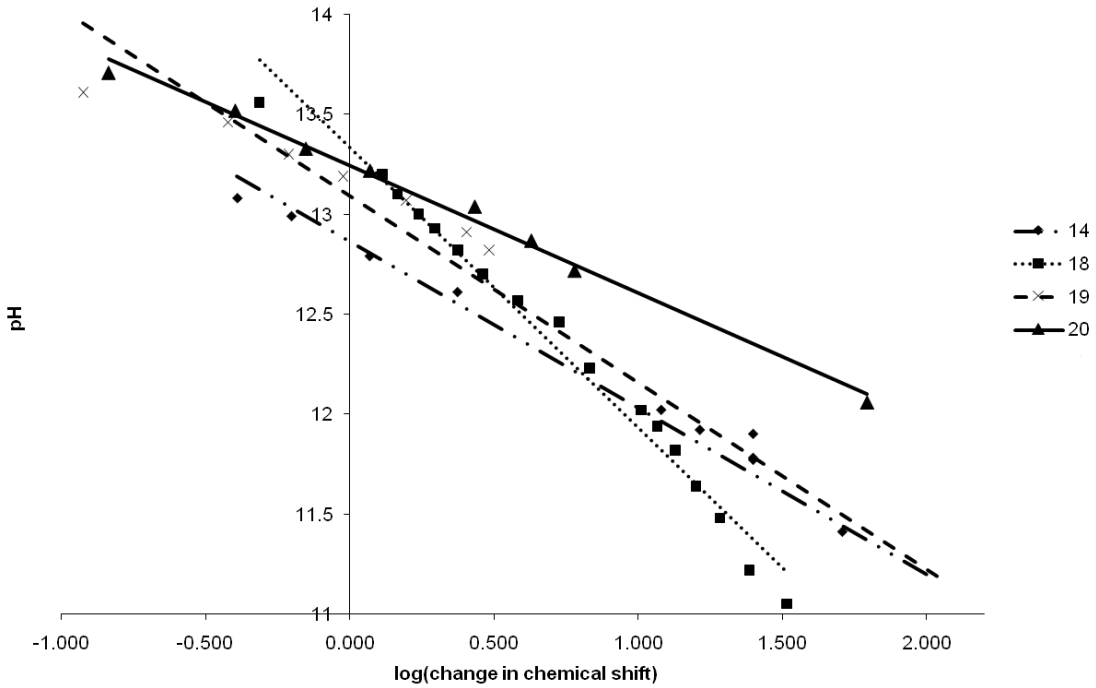


Figure 9. **Hill plots to calculate SMoC pKa values.** The log of the ratio of uncharged against charged species expressed in terms of chemical shift plotted against pH of the solution to find the pKa value as the y-intercept.

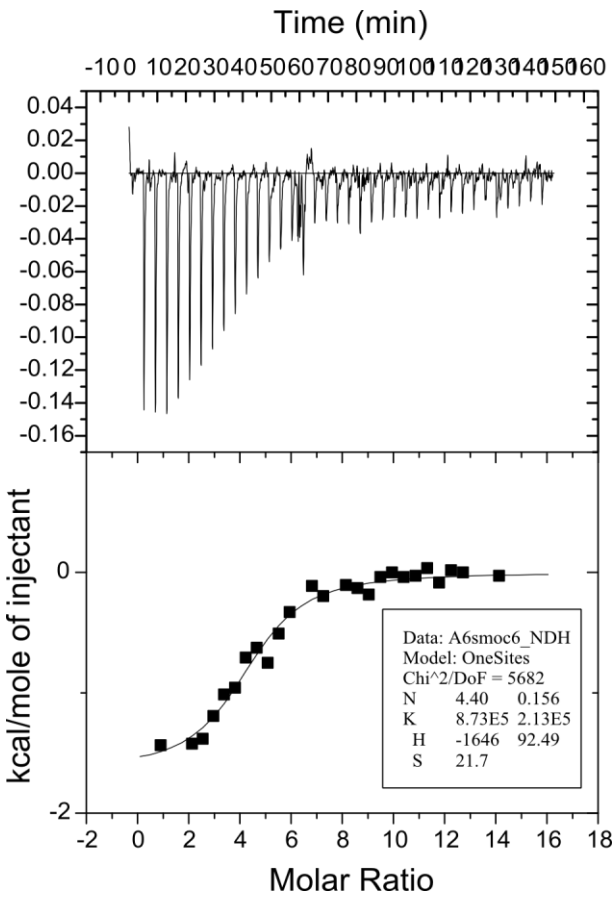


Figure 10. **6G-SMoC ITC binding curve.** Above: ITC data for titration of 6G-SMoC against human GAPD siRNA. Each spike represents the heat evolved upon a single 8 µl injection of 250 µM 6G-SMoC, **24** into the calorimeter cell containing 3.5 µM GAPD siRNA. Below: The Origin software was used to fit the ITC data to a binding curve and calculate the binding parameters.

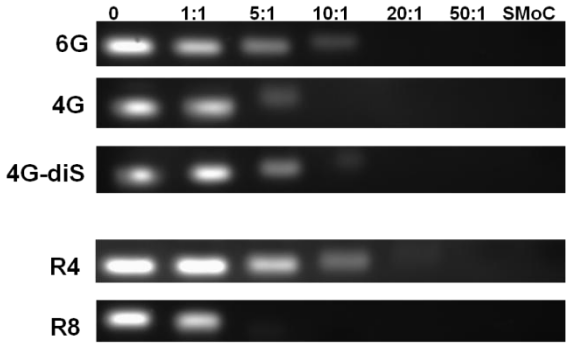


Figure 11. siRNA gel shift assay. GAPDH siRNA was mixed with SMoC compounds or a polyarginine peptide (R4, R8) at the ratios shown with a running buffer containing ethidium bromide and run on a 1% agarose gel. The gels were visualised under UV illumination. SMoCs alone gave no signal (not shown).

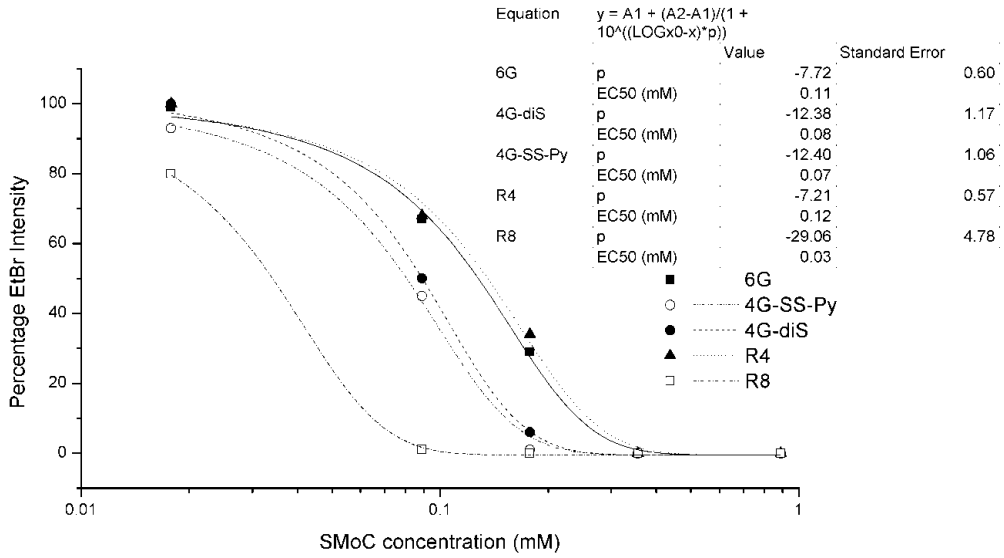


Figure 12. **Binding curves derived from gel shift assay data.** The EtBr bands were quantified and the band intensity plotted to produce binding curves for each of the compounds with siRNA. The EC_{50} and Hill coefficient (p) values of each compound are shown, giving a relative indication of the binding affinity of each compound to siRNA.

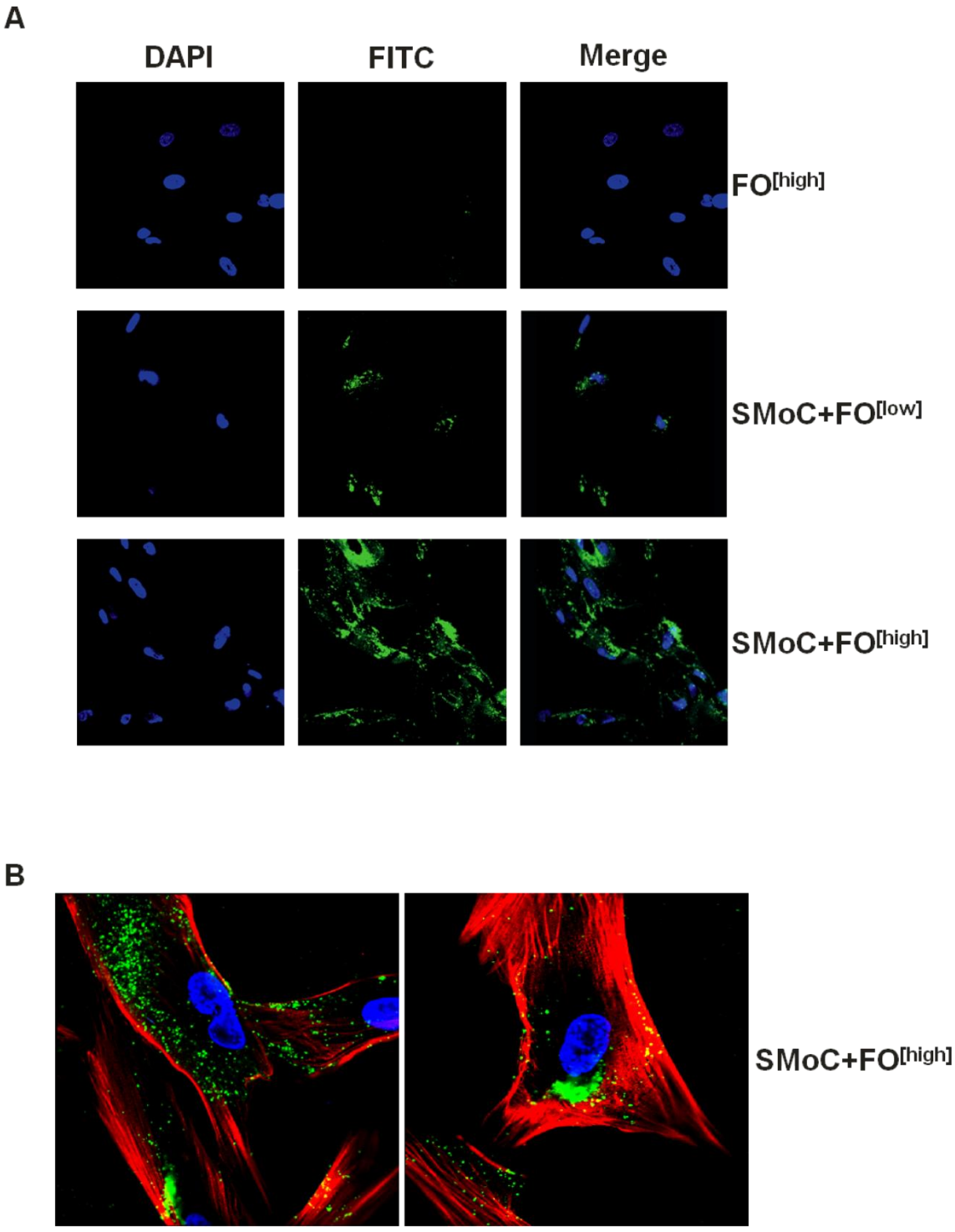


Figure 13. Intracellular delivery of fluorochrome-labelled siRNA (FO) by 4G-SMoC. (**A)** IMR-90 human diploid fibroblasts were cultured on coverslips and treated for 24 hours with 100 nM FO ($FO^{[high]}$) or complexes of **10**, 4G-SMoC (40 µg/ml, 35 µM) and either 10 nM FO ($FO^{[low]}$) or 100 nM FO ($FO^{[high]}$). (**B)** IMR-90 cells treated with 100 nM FO ($FO^{[high]}$) in complex with **10**, SMoC and stained with DAPI and AlexaFluor 546 Phalloidin after fixation with paraformaldehyde.

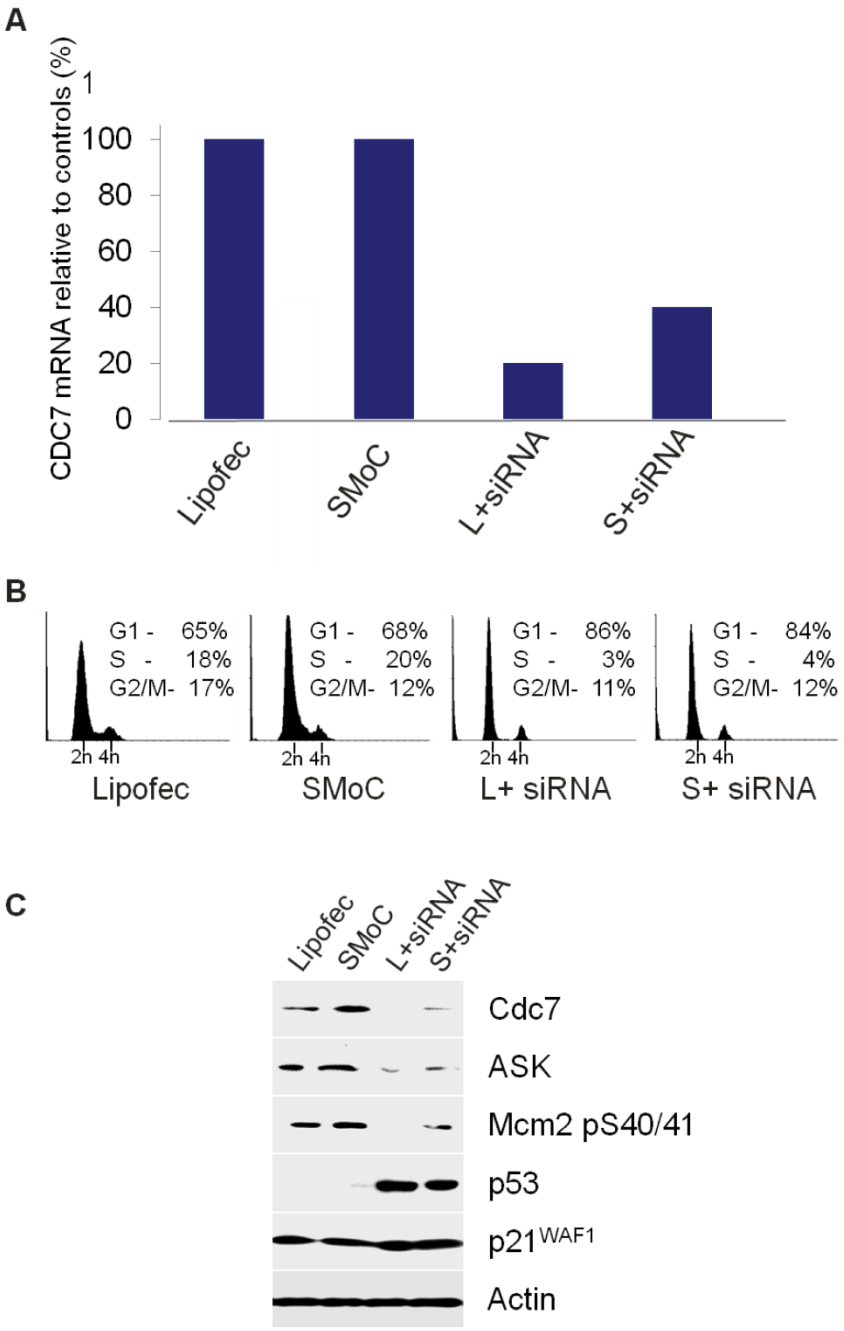


Figure 14. CDC7 knock down phenotype in IMR90 cells after SMoC-mediated uptake of CDC7-siRNA. (**A**) IMR-90 fibroblasts were treated for 72 hours with complexes of either Lipofectamine 2000 or 4G-SMoC **10**, with CDC7-siRNA. Lipofectamine 2000 and **10** alone were used as controls. Knock down of CDC7 transcript levels was confirmed by qRT-PCR relative to Lipofectamine only and **10**, only treatment. (**B**) DNA content of Lipofectamine-treated cells (Lipofec), **10** - treated cells, and cells treated with complexes of Lipofectamine and CDC7-siRNA (L+siRNA) or **10** and CDC7-siRNA (**10**+siRNA). (**C**) Whole cell extracts prepared from Lipofec, **10**-treated, L+siRNA and **10**+siRNA cells were analysed by immunoblotting with the indicated antibodies (β-actin—loading control).

Bibliography

1. Guo J, Fisher KA, Darcy R, Cryan JF, O'Driscoll C (2010) Therapeutic targeting in the silent era: advances in non-viral siRNA delivery. *Molecular Biosystems;***6**: 1143-61.

2. Whitehead K, Langer R, Anderson D (2009) Knocking down barriers: advances in siRNA delivery. *Nature Reviews Drug Discovery;***8**: 129-38.

3. Sontheimer EJ (2005) Assembly and function of RNA silencing complexes. *Nature Reviews Molecular Cell Biology;***6**: 127-38.

4. Li W, Szoka FC (2007) Lipid-based nanoparticles for nucleic acid delivery. *Pharm Res;***24**: 438-49.

5. Rozema DB, Lewis DL, Wakefield DH, Wong SC, Klein JJ, Roesch PL, et al. (2007) Dynamic PolyConjugates for targeted in vivo delivery of siRNA to hepatocytes. *Proceedings of the National Academy of Sciences of the United States of America;***104**: 12982-7.

6. Kim SW, Kim NY, Choi YB, Park SH, Yang JM, Shin S (2010) RNA interference in vitro and in vivo using an arginine peptide/siRNA complex system. *Journal of Controlled Release;***143**: 335-43.

7. Konate K, Crombez L, Deshayes S, Decaffmeyer M, Thomas A, Brasseur R, et al. (2010) Insight into the cellular uptake mechanism of a secondary amphipathic cell-penetrating peptide for siRNA delivery. *Biochemistry;***49**: 3393-402.

8. Eguchi A, Dowdy SF (2009) siRNA delivery using peptide transduction domains. *Trends Pharmacol Sci;***30**: 341-5.

9. Veldhoen S, Laufer SD, Trampe A, Restle T (2006) Cellular delivery of small interfering RNA by a non-covalently attached cell-penetrating peptide: quantitative analysis of uptake and biological effect. *Nucleic Acids Res;***34**: 6561-73.

10. Crombez L, Aldrian-Herrada G, Konate K, Nguyen QN, McMaster GK, Brasseur R, et al. (2009) A new potent secondary amphipathic cell-penetrating peptide for siRNA delivery into mammalian cells. *Molecular Therapy;***17**: 95-103.

11. Gump JM, June RK, Dowdy SF (2010) Revised role of glycosaminoglycans in TAT protein transduction domain-mediated cellular transduction. *Journal of Biological Chemistry;***285**: 1500-7.

12. Jiao CY, Delaroche D, Burlina F, Alves ID, Chassaing G, Sagan S (2009) Translocation and Endocytosis for Cell-penetrating Peptide Internalization. *J Biol Chem;***284**: 33957-65.

13. El-Sayed A, Khalil IA, Kogure K, Futaki S, Harashima H (2008) Octaarginine- and Octalysine-modified Nanoparticles Have Different Modes of Endosomal Escape. *J Biol Chem;***283**: 23450-61.

14. Chao T-Y, Raines RT (2011) Mechanism of Ribonuclease A Endocytosis: Analogies to Cell-Penetrating Peptides. *Biochemistry;*null-null.

15. Duchardt F, Fotin-Mleczek M, Schwarz H, Fischer R, Brock R (2007) A Comprehensive Model for the Cellular Uptake of Cationic Cell-penetrating Peptides. *Traffic;***8**: 848-66.

16. Burley SK, Petsko GA (1986) Amino-aromatic interactions in proteins. *FEBS Lett;***203**: 139-43.

17. Gallivan JP, Dougherty DA (1999) Cation-pi interactions in structural biology. *Proceedings of the National Academy of Sciences of the United States of America;***96**: 9459-64.

18. Rebstock AS, Visintin C, Leo E, Posada CG, Kingsbury SR, Williams GH, et al. (2008) Modular assembly using sequential palladium coupling gives easy access to the SMoC class of cellular transporters. *ChemBioChem;***9**: 1787-96.

19. Okuyama M, Laman H, Kingsbury SR, Visintin C, Leo E, Eward KL, et al. (2007) Small-molecule mimics of an [alpha]-helix for efficient transport of proteins into cells. *Nat Methods;***4**: 153-9.

20. Lundberg P, El Andaloussi S, Sutlu T, Johansson H, Langel U (2007) Delivery of short interfering RNA using endosomolytic cell-penetrating peptides. *FASEB J;***21**: 2664-71.

21. Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins;***47**: 393-402.

22. Wang JM, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem;***21**: 1049-74.

23. Kingsbury SR, Loddo M, Fanshawe T, Obermann EC, Prevost AT, Stoeber K, et al. (2005) Repression of DNA replication licensing in quiescence is independent of geminin and may define the cell cycle state of progenitor cells. *Exp Cell Res;***309**: 56-67.

24. Tudzarova S, Trotter MWB, Wollenschlaeger A, Mulvey C, Godovac-Zimmermann J, Williams GH, et al. (2010) Molecular architecture of the DNA replication origin activation checkpoint. *EMBO J;***29**: 3381-94.

25. Wang XJ, Sun XF, Zhang L, Xu YB, Krishnamurthy D, Senanayake CH (2006) Noncryogenic I/Br-Mg exchange of aromatic halides bearing sensitive functional groups using i-PrMgCl-bis 2-(N,N-dimethylamino)ethyl ether complexes. *Org Lett;***8**: 305-7.

26. Molander GA, Figueroa R (2005) Organotrifluoroborates: Expanding organoboron chemistry. *Aldrichimica Acta;***38**: 49-56.

27. Molander GA, Ellis N (2007) Organotrifluoroborates: Protected boronic acids that expand the versatility of the Suzuki coupling reaction. *Acc Chem Res;***40**: 275-86.

28. Ducatel H, Van Nhien AN, Postel D (2008) Novel glycine like amino acids from glyco-[alpha]-aminonitriles as building blocks for peptide synthesis. *Tetrahedron: Asymmetry;***19**: 67-81.

29. Feldman KS, Ngernmeesri P (2005) Dragmacidin E synthesis studies. Preparation of a model cycloheptannelated indole fragment. *Org Lett;***7**: 5449-52.

30. Feichtinger K, Sings HL, Baker TJ, Matthews K, Goodman M (1998) Triurethane-Protected Guanidines and Triflyldiurethane-Protected Guanidines: New Reagents for Guanidinylation Reactions. *The Journal of Organic Chemistry;***63**: 8432-9.

31. Acharya P, Plashkevych O, Morita C, Yamada S, Chattopadhyaya J (2003) A repertoire of pyridinium-phenyl-methyl cross-talk through a cascade of intramolecular electrostatic interactions. *J Org Chem;***68**: 1529-38.

32. Oganesyan A, Cruz IA, Amador RB, Sorto NA, Lozano J, Godinez CE, et al. (2007) High yield selective acylation of polyamines: proton as protecting group. *Organic Letters;***9**: 4967-70.

33. Atkins P, de Paula J (2002) *Atkins' Physical Chemistry*. 7 edn: Oxford University Press.

34. Kumaki Y, Nitta K, Hikichi K, Matsumoto T, Matsushima N (2004) Side Chain-Side Chain Interactions of Arginine with Tyrosine and Aspartic Acid in Arg/Gly/Tyr-Rich Domains within Plant Glycine-Rich RNA Binding Proteins. *J Biochem;***136**: 29-37.

35. Swords R, Mahalingam D, O'Dwyer M, Santocanale C, Kelly K, Carew J, et al. (2010) Cdc7 kinase - A new target for drug development. *Eur J Cancer;***46**: 33-40.

36. Won Y-W, Kim HA, Lee M, Kim Y-H (2010) Reducible Poly(oligo-D-arginine) for Enhanced Gene Expression in Mouse Lung by Intratracheal Injection. *Mol Ther;***18**: 734-42.

37. Aubry S, Aussedat B, Delaroche D, Jiao CY, Bolbach G, Lavielle S, et al. (2010) MALDI-TOF mass spectrometry: A powerful tool to study the internalization of cell-penetrating peptides. *Biochim Biophys Acta-Biomembr;***1798**: 2182-9.

38. Eguchi A, Meade BR, Chang YC, Fredrickson CT, Willert K, Puri N, et al. (2009) Efficient siRNA delivery into primary cells by a peptide transduction domain-dsRNA binding domain fusion protein. *Nat Biotechnol;***27**: 567-U110.

39. Lensink MF, Christiaens B, Vandekerckhove J, Prochiantz A, Rosseneu M (2005) Penetratin-membrane association: W48/R52/W56 shield the peptide from the aqueous phase. *Biophys J;***88**: 939-52.